Molecular Techniques in Crop Improvement

Book · January 2010

DOI: 10.1007/978-90-481-2967-6



Shri Mohan Jain
University of Helsinki

Molecular Techniques in Crop Improvement

S. Mohan Jain • D.S. Brar
Editors

Molecular Techniques in Crop Improvement

2nd Edition

Springer

Editors
S. Mohan Jain
Department of Applied Biology
University of Helsinki
Finland
mohan.jain@helsinki.fi

D.S. Brar
International Rice Research
Institute (IRRI)
Social Sciences Division
Metro Manila
Philippines
d.brar@cgiar.org

ISBN 978-90-481-2966-9 e-ISBN 978-90-481-2967-6
DOI 10.1007/978-90-481-2967-6
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2009929346

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Most of the plant breeding programs aim to increase yield, disease and insect resistance, abiotic stress tolerance and to improve quality characteristics. The value of new plant breeding products and varieties in increasing food production has been demonstrated time and again. To meet growing need of ever increasing human population, we need to enhance food production for sustaining food supply. Furthermore, several biotic and abiotic stresses continue to threaten crop productivity. Moreover with urbanization, land for cultivation is shrinking and several environment concerns involving excessive use of fertilizers and agro-chemicals, soil and water pollution including water scarcity are key issues in increasing crop productivity and food sustainability. Plant breeders therefore, has the major challenge how to increase crop productivity with limited land, limited water, limited chemicals and limited labour particularly in the context of global climate changes. In the genomics era, advances in molecular biology have opened new opportunities to accelerate plant breeding processes and in overcoming some of the above constraints limiting crop productivity. Molecular markers have become important tools in the hands of plant breeders in marker assisted breeding and for enhancing the selection efficiency for various agronomic traits in precision 'plant' breeding. The isolation, cloning and moving of genes from diverse biological sources into plant genomes holds promise to broaden the gene pool of crops and develop new plant varieties for specific traits that determine yield, quality, and resistance to biotic and abiotic stresses. New genomics tools will be of great value to support conventional breeding for sustainable food production especially under the climate change and meet demand of ever growing human population.

The first edition of this book, "Molecular techniques in crop improvement" published in 2002 covered various topics related to molecular markers and their application in plant breeding. Since then, major advances have been made in molecular tagging of genes/QTLs governing complex agronomic traits, identification of candidate genes and in applying marker assisted breeding for tolerance to biotic and abiotic stresses and quality traits. Recent advances in transgenic technologies, genome sequencing and functional genomics offer tremendous opportunities to support plant breeding programs. Therefore, we are encouraged to cover recent advances and come out with the second edition of this book. In this edition we have included 31 chapters, which are divided into four parts. Part I is on Plant

Breeding in the genomic era and has four chapters. Part II deals with Molecular Markers and their application and contains six chapters. Eleven chapters dealing with different aspects of Genomics are covered in Part III. The remaining ten chapters dealing with Transgenic Technologies are dealt in Part IV. Some of the major topics covered in this book are on: QTL analysis, comparative genomics, functional genomics, bioinformatics, gene-based marker systems, automation of DNA marker analysis, molecular markers for abiotic and biotic stresses as well as for germ plasm conservation, gene pyramiding, gene stacking, gene silencing, TILLING, CISGENESIS, microarray, metablomics, proteomics, transcriptomics, microRNAs, marker-free transformation, Plant RNAi, and floriculture genetic engineering. All manuscripts were peer reviewed and revised accordingly. We are thankful to all reviewers for sparing precious time to review manuscripts, and that certainly helped to improve their quality.

This edition will certainly benefit plant breeders, biotechnologists, and molecular biologists, and an excellent source to provide advanced knowledge of molecular biology to post graduate students involved in biotechnology and plant breeding research.

We take this opportunity to express our gratitude to Springer publisher for giving us this opportunity to bring out the second edition of this book.

S. Mohan Jain
D.S. Brar

Contents

Part I
Plant Breeding in the Genomics Era

Chapter 1
QTL Analysis in Plant Breeding

Maria J. Asins, Guillermo P. Bernet, Irene Villalta, and Emilio A. Carbonell

Abstract To maintain the quality of life on earth, agriculture has to colonize marginal areas and decrease its dependence on pesticides, herbicides, fertilizers and water. Plant breeding should integrate the latest innovations in biology and genetics to better face this challenge. Quantitative trait loci (QTL) analysis allows the location and effect-estimation of the genetic elements controlling any trait by the joint study of segregation of marker genotypes and of phenotypic values of individuals or lines. QTL analysis is now seen as a procedure to fill the gap between "omics" and the field. This chapter is focused on recent advances of three applications of QTL analysis in plants: (1) the genetic integration of agronomical, physiological and gene expression related traits (the scientific value of QTL analysis); (2) Marker-assisted selection (MAS) in breeding programs and (3) the utilization of wild germplasm to improve quantitative traits with breeding tomato for salt tolerance as an example.

1.1 Introduction

The general goal of plant breeding is the improvement of plants for human benefit, fulfilling the needs of both producers and consumers. However, the exponential growth of human population, the effects of the global warming on the crop environments and the need to stop the negative impact of agriculture on the ecosystems, are demanding new and urgent specific goals (of particular social value) to maintain the quality of life on earth. In short, a more sustainable and environmental friendly agriculture that increases crop yield taking into account the tolerance to biotic and abiotic stresses and the efficiency of water and nutrient uptake. This will allow colonizing marginal areas for agriculture and decreasing its dependence on pesticides, herbicides, fertilizers and water. The great scientific advances of twenty-first century should help to pace these new challenges.

M.J. Asins (✉), G.P. Bernet, I. Villalta, and E.A. Carbonell
IVIA, Apdo. Oficial, 46113, Moncada, Valencia, Spain
e-mail: mjasins@ivia.es

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_1, © Springer Science+Business Media B.V. 2010

Plant breeding has a long history of integrating the latest innovations in biology and genetics to enhance crop improvement (Moose and Mumm 2008). The large advance in science has stimulated shifts in funding at public institutions to enhance intellectual capacity and infrastructure for molecular genetics and genomics, often, ironically, at the expense of conventional plant breeding (Knight 2003; Brummer 2004). Likely linked to this, plant breeding has often moved from public institutions to commercial companies. There is a need of research linking molecular methods with breeding objectives to fully realize the potential of recent advances in biotechnology and genomics (Guimarâes and Kueneman 2006; National Research Council 2008)

Nowadays, there are two different strategies for molecular improvement depending on the origin of genetic variability: hybridization of genotypes (the classical one), and genetic transformation and TILLING (Targeting Induced Local Lesions in Genomes, Slade et al. 2005) of an elite genotype. The former uses natural genetic variability, and allows the obtaining of multiple combinations integrated by multiple variants of (usually) small effect each. The latter obtains new forms yielding few variants that contain, in the case of transformation, one or few copies of a functionally important gene or DNA fragment (supposedly of large effect on the phenotype) and in the case of TILLING, a mutated allele of the gene of interest. Both alternatives manage very different amounts of genetic variability (large and small, respectively) and provide complementary information (from forward and reverse genetics, respectively). Although the latter, if successful, might seem more attractive due to the rapid genetic gain in the trait of interest, breeding programs must cope with multiple agronomic traits (yield, quality, timing, resistance …), most of them controlled by multiple genes (some of them linked or with pleiotropic effects) and interactions (with other genes and the environment). Therefore, at least the knowledge of the position of the genes controlling these traits is relevant.

Only by the joint analysis of segregation of marker genotypes and of phenotypic values of individuals or lines, it is possible to detect and locate loci affecting quantitative traits ("quantitative trait loci" or "QTLs"). QTLs are, a priori, difficult to identify due to the lack of discrete phenotypic segregation and because the phenotypic effect of each gene associated with a complex trait is relatively small. QTL analysis in segregating populations, involves selecting and hybridizing parental lines that differ in one or more quantitative traits and analyzing the progeny in order to link the QTL to known DNA markers. A breeder can use this knowledge to advantage, for instance by using indirect selection. When selection is (partly) based on genetic information retrieved through the application of molecular markers this is called marker-assisted selection (MAS). It can be employed to enhance plant breeding efforts and to speed up the creation of cultivars. Also, it unveils masked, interesting wild alleles and makes for an easier introduction of genetic material from related and unrelated wild species, without the drawbacks ("linkage drag") that are associated with the introduction of "wild genes" through conventional methods, facilitating germplasm enhancement and pre-breeding.

QTL analysis is usually seen as the methodology of choice to study quantitative traits, which show continuous variation and are theoretically controlled by many genes. But, a priori, any trait (not only the quantitative traits) can be controlled by more than one gene what makes QTL analysis an universal tool to obtain information

on the number and position of genes controlling it. We may detect just one QTL in a given segregating population but by using this genomic scanning tool, other hypothesis (more than one locus) can be tested, too.

QTL analysis has clear limitations that have been reviewed by several authors (Doerge and Rebai 1996; Beavis 1994; Kearsey and Farquhar 1998; Doerge 2002; Asins 2002; Melchinger et al. 2004; Holland 2007). A major problem, pointed by Göring et al. (2001), come from the fact that the chromosomal position and genotype-phenotype relationship of a locus cannot both be reliably estimated by the use of a single data set of currently realistic size, at least for loci of small effect size. An approach put forward by Sen and Churchill (2001) breaks the QTL problem into two distinct parts: the relationship between the QTL and the quantitative trait, and the location of the QTL. So the initial focus is placed on estimation of the unknown QTL genotypes and then on allowing the search for different models and their comparisons with the information gained from completing the QTL genotype information (Doerge 2002). Other improvements concerning the experimental design (segregating populations from multiple parents and environments) and the development of new statistical methodologies for locating multiple QTLs (reviewed by Zou and Zeng 2008) will enhance our capability of detecting more QTLs per trait, including those of small effect. Besides, recent advances in statistical methods to control the QTL false-discovery rate should provide a better balance between declaring too many false-positive QTL and sacrificing power to detect those that have smaller effects (Benjamini and Yekutieli 2005).

Advances in the application of QTL analysis to plants are also becoming evident.

The most obvious are marker-assisted selection (MAS) in breeding and pre-breeding and QTL cloning. Other areas where QTL analysis is contributing decisively are: (1) understanding of complex but agronomically important traits such as plant-pathogen interaction and adaptability to marginal areas (genotype-environment interaction); (2) plant genomics, providing information on candidate polymorphisms by connecting the QTLs that control phenotypes, metabolites, proteins, genes and regulatory elements; and (3) germplasm enhancement allowing its efficient utilization in pre-breeding through genotyping for candidate polymorphisms. All of them, and particularly MAS, have been a major impetus to quantitative genetics research and breeding.

The present chapter focuses on recent advances of three applications of QTL analysis in plants: (1), the genetic integration of agronomical, physiological and gene expression traits (the scientific value of QTL analysis) (2) MAS in breeding programs and (3) utilization of wild germplasm to improve quantitative traits, with breeding tomato for salt tolerance as an example.

1.2 The Scientific Value of QTL Analysis

QTL analysis has become the methodology of choice for the genetic dissection of any trait in plant and animal species. QTL cloning is usually the next objective. The list of cloned QTLs has been expanded in the past few years (Paran and Zamir

2003; Morgante and Salamini 2003; Salvi and Tuberosa 2005; Price 2006). A cloning approach starts by searching for functional candidates, i.e. genes that have been shown, or are suspected, to have a functional role in the phenotype of interest. Then, the most promising candidates are selected from a large number of functional candidate genes, by testing their linkage to the QTLs for the trait of interest, thereby identifying positional candidates (genes co-locating with QTLs) (Pajerowska et al. 2005). This approach is particularly useful for discarding candidates as shown by Villalta et al. (2008) using sequences coding for proteins involved in Na^+ transport as functional candidates for salt tolerance in terms of fruit yield. Most QTLs have been cloned using a positional cloning approach. It requires to assign a QTL to the shortest possible genetic interval (QTL fine mapping) and to identify the corresponding interval on the DNA sequence (QTL physical mapping) where candidate genes are selected for evaluation (Salvi and Tuberosa 2005). The availability of a full-genome sequence is a helpful tool in filtering through genes in the interval, because the examination of annotation can often suggest which of the genes in the QTL interval might be likely candidates (Hansen et al. 2008). The candidate approach somehow bypasses the tedious procedures of positional cloning. Besides, although the size of the QTL confidence interval might be a limitation because of the large number of candidate genes included in a minimum of 10 cM, the position of the maximum significance for the QTL detection might be a good indication of the position of the responsible gene (Price 2006). A possible complication may arise when functional candidates are organized in clusters, such as defense-related genes and genes participating in cell-wall integrity frequently are (Vergne et al. 2008). Interestingly, these authors found that these clusters are not always co-regulated and individual paralogs can show specific expression patterns. Therefore, the identification of suitable candidate genes and the elucidation of their function can be facilitated by combining different approaches and high-throughput platforms applied to the target crop and/or to model species.

Although gene-expression analysis and its results provide valuable information towards understanding gene function, these results are not helpful in relating gene function to complex traits, because they do not address anything other than the relative change in gene expression across treatments (Doerge 2002). A major step forward in QTL cloning has occurred via the application of microarray technology to obtain genome-wide expression profiling from individuals in a RIL population. This enables the mapping of QTLs controlling the transcript level for each gene (expression QTLs, eQTLs) and, thereby, the study of the relationship between genome and transcriptome. These eQTLs can be utilized to search for associations between gene expression polymorphisms and a phenotypic QTL to identify candidate genes controlling phenotypic variation for an agronomic trait (Hansen et al. 2008). Moreover, if the expression level of single genes involved in a process (metabolic pathway, development, defense response …) is converted into a common measure (z scaling or mean shifting) then, their QTL analysis will reveal network eQTLs; i.e. QTLs controlling gene expression networks or polymorphisms upstream in the process or pathway (Kliebenstein et al. 2006). Results from eQTL analysis can also be used to generate networks using either correlation of expression

patterns or colocalization of eQTL positions to identify clusters or networks of genes (Lan et al. 2006). Then, trans eQTL can be identified for these novel networks, and subsequently these genetic loci can be searched for regulatory genes containing *cis* eQTLs (Sun et al. 2007). This "systems biology" approach to natural variation using the same fixed segregant populations may eventually allow to link metabolites in a biosynthetic network with network eQTLs for the biosynthetic genes towards the understanding of the molecular basis of phenotypic QTLs (Hansen et al. 2008; Lisec et al. 2008).

Among the different platforms available for mass-scale profiling at the transcriptome, microarrays have been frequently used to study the changes in gene expression elicited by exposure to biotic or abiotic stresses. Because the functional basis of a number of cloned plant QTLs have been found to regulate differences in the level of gene expression (Salvi and Tuberosa 2005), QTL cloning may in some cases be facilitated through a direct profiling approach to suitable genetic materials (Wayne and McIntyre 2002; Hazen et al. 2003). In this context, the QTL analysis of the level of gene expression (eQTL analysis) is a promising application of transcriptome analysis. More recently, cDNA-AFLPs have been used as an alternative to microarrays to identify eQTLs in Arabidopsis (Vuylsteke et al. 2006). Important drawbacks of the latter strategy are the limited coverage of the transcriptome and the identification of differential genes which requires purification and sequencing of individual AFLP fragments. The main drawback of the former strategy is the high cost for profiling RNA samples of an entire population but a benefit of eQTL analysis is that the same arrays can be simultaneously used for genetic mapping and expression quantification (Luo et al. 2007; West et al. 2006). With recent advances in genomics, the bottleneck in QTL and marker-trait association analysis will be the phenotyping and not the genotyping (Melchinger et al. 2004; Xu and Crouch 2008).

Since the number of publications reporting QTL analysis is continuously growing, it would be desirable to integrate information from different QTL mapping experiments to get a more realistic and complete view of the genetic control of traits. Some agronomic traits are so important that numerous QTL studies using multiple segregating populations have been published. One early example integrating QTL analyses of fruit size and shape in tomato was reported by Grandillo et al. (1999) giving a wide view of the location and consistency (and/contribution) of the approximately 28 QTLs that control fruit weight. This information was important to focus the subsequent QTL cloning experiments of the group. Another approach to enhance the inference space of QTL studies is meta-analysis, whereby the results of multiple studies are summarized and evaluated statistically (Chardon et al. 2004; Wisser et al. 2006; Ballini et al. 2008). As Holland (2007) has pointed out, these studies provide useful summaries of genetic architecture; however, they are complicated by ascertainment bias, by differences in experimental techniques, statistical analysis and environments (grennhouse, field …). Besides, they implicitly assume QTL additivity while epistasis, particularly, less than additive epistasis has been reported for some traits (in tomato, for instance, Lecomte et al. 2004; Causse et al. 2007; Estañ et al. 2008).

An example of further level of information integration has been provided by Vergne et al. (2008) who cross-referencing defense gene expression (more than 2,500 genes corresponding to the rice defense arsenal) and resistance QTL information have identified some candidates (overall pathogenesis-related genes and disease regulators). Obviously, this level of integration is only possible on model, fully sequenced, crops where the genetic map is anchored to a physical map, facilitating the assignment of genes, by function and position, to QTLs. After this, linkage disequilibrium studies are expected to facilitate validation and map-based cloning of genes (Gupta et al. 2005; Salvi and Tuberosa 2005).

Therefore, QTL analysis allows connecting phenotypic variation with allelic variation at a position in the genome, and then through fine mapping and/or linkage disequilibrium or association mapping to pinpoint the "functional" nucleotide polymorphism or haplotype best explaining germplasm variability (Veyrieras et al. 2007). Association mapping provides an alternate route to identifying or validate QTL that have effects across a broader spectrum of germplasm. It requires identifying, within a set of genotypes such as germplasm accessions and cultivated varieties, a statitistical association between allelic variants at marker or candidate loci and the mean of the analyzed trait. A very simple example of marker validation has been recently reported by Bernet et al. (2008). They used seven markers linked to QTLs involved in *Citrus tristeza virus* (CTV) accumulation, leaf miner resistance and apomictic reproduction to characterize 64 sour orange (*C. aurantium* L.) accessions from three national collections in order to identify a representative core in which the resistance behavior against two CTV isolates was studied. Most of the *C. aurantium* accessions fell into three main groups based on three multilocus genotypes. Only those selected accessions from the major group presented CTV resistance. When the haplotype diversity at three CTV accumulation QTL-linked markers that defined those groups was further studied by sequence analysis, only allelic diversity at one of them tightly linked to *Ctv-R*$_2$ in *Poncirus trifoliata* (L.) Raf., matched the plant-CTV interaction types reported among *Poncirus* and *Citrus* species.

Thus, finally, the huge scientific advances due to QTL mapping and "omics" resources, after validation are expected to be translated into heritable assays that are more effective and efficient than phenotypic assays to enhance germplasm and plant breeding (Beavis et al. 2007).

1.3 QTL Analysis for MAS

Breeding objectives always involve consideration of multiple traits, even in situations where output of a single feature is dominant. While animal breeders have formalized the procedure of multiple trait selection, commercial plant breeders use non-formal ways of combining selection pressure on various traits, which are not published (Sölkner et al. 2008). The longer generation intervals and, more importantly, the lower reproductive rates of animals emphasize procedures of simultaneous selection whereas crops are developed in multiple stages, emphasizing different

traits at different generations. Thus, selection for traits with high heritability is usually performed at early generations. If a kind of pre-selection based on QTL linked markers might be carried for one or more traits at seedling stage, fewer generations would be needed. Long juvenility species would greatly benefit from this. However, to ensure the success of MAS, QTL mapping results, particularly QTL estimates, must be further investigated by cross validation and biometrical methods (Melchinger et al. 2004), even including Bayesian approaches (Meuwissen et al. 2001). Even so, there will always be QTLs of minor effects (and epistatically masked QTLs) that will be missed, making selection of favorable QTL alleles alone unable to reach the full genetic potential and the ultimate limits to selection (Kearsy et al. 2003).

There are two important genetic factors, epistasis and genotype by environment interaction (G × E) that limits our understanding of agronomic traits. In this sense, the use of marker-assisted selection (MAS) can be inefficient if the effects of G × E and epistasis cannot be anticipated (Openshaw and Frascaroli 1997; Moreau et al. 2004). Moreover, Bernardo and Charcosset (2006) showed that poor estimates of allele effects within the selected population, even when genes are known, reduced MAS efficiency. Therefore, when G × E and epistasis are important we should regularly re-estimate QTL effects within the breeding program (Podlich et al. 2004).

Significant G × E makes statistically impossible to interpret G and E main effects and to predict genotypic performance across changing environments. In its simplest form, G × E may rise from heterogeneous genetic variances among environments. Thus, genetic variance at one QTL may be sufficiently large in one environment, in comparison to non-genetic effects, to allow its detection here, but not in other environment (salinity, for instance). If it happens in one or more QTLs for the same environment, it will cause just a scaling effect but it will not alter the ranking of genotypes. But, if the variance heterogeneity affects different QTLs depending on the environments, a re-ranking of genotypes will happen which is a major concern to breeders, particularly those devoted to increase crop adaptability to abiotic stresses.

To determine genetic factors responsible for G × E, agronomic data must be collected on a mapping population in multiple environment trials. Then, comparison of QTL detection across environments is carried out by analysis of variance to test marker locus × environment interactions (Sari-Gorla et al. 1997; Villalta et al. 2007). Quantitative trait loci by environment interaction can also be evaluated by the regression of marker genotype mean on an environmental index to discern if the linear regression coefficients are significantly different (Campbell et al. 2003). QTL mapping results (e.g. Patterson et al. 1991; Stuber et al. 1992; Lu et al. 1996; Monforte et al. 1997b; Villalta et al. 2007) have shown that some QTLs can be detected in all tested environments while others are detected in some of them. However, in the absence of G × E, a QTL may be detected in one environment but not in others due to sampling or experimental error. On the other hand, G × E may exist even when QTL are detected in multiple environments (Yan et al. 1999). Noteworthy, Villalta et al. (2008) found that the position of maximum LOD for a

leaf area QTL (and also for dried leaf weight) on *Solanum* chromosome 5, changed few cM when control and salinity growing conditions were compared. Only QTLs detected under control condition showed significant G × E interaction. Is it a matter of random errors or the presence of two QTLs in tandem, one conditioning the trait just under control and the other under both control and salinity conditions? Unfortunately, random errors cannot be discarded given the overlapping of confidence intervals.

Due to the cost of utilizing several QTLs, only markers that are tightly linked to no more than three QTLs are typically used for MAS (Ribaut and Betran 1999), although there have been reports of up to five QTLs being introgressed via MAS (Lecomte et al. 2004). The cost of using MAS compared to conventional plant breeding varies considerably between studies and need to be considered in a case-by-case basis. And, as Collard et al. (2005) have pointed out these studies did not include the large initial cost in their development. An estimate for the cost to develop a single marker was 51,140 € (Langridge et al. 2001). Model crops, where genomic tools and most QTL analysis are being developed have a clear advantage with regard to applications of QTL mapping. Fortunately, comparative mapping could be used to infer QTL position between related species what could be particularly important in “orphan” (or neglected) crops.

Markers tightly linked to loci controlling difficult, laborious and expensive but breeding-targeted traits are valuable tools to assist selection at seedling stage during the breeding program of a long-lasting juvenility crop or forest species. Thus, MAS have been most frequently used (or reported) to discard putative disease susceptible plants in the early generations and to introgress disease resistance genes into high-quality well adapted elite cultivars. The greatest efficiency of MAS is in early generations due to the increasing probability of recombination between the marker and QTL; the major disadvantage is the cost of genotyping a large number of plants given that it is applied at the beginning of the program (Collard and Mackill 2008). The continuously growing marker technology and availability of high-throughput equipment for DNA extraction and genotyping (for a review see Gupta et al. 2008) is expected to make MAS less expensive in the future.

A successful strategy to minimize the failure of marker-trait association is to maximize the efficiency of QTL mapping in tandem with the MAS process itself (Young 1999). Additionally, to cover a wider range of allelic diversity, multiparental populations (like those commonly used by breeders) have shown to be better than biparental populations. Thus, Blanc et al. (2006) have recently shown by simulation that addressing a broader diversity, multiparental designs increase the power of QTL detection, which reinforces their superiority over biparental designs for MAS. As Melchinger et al. (2004) concluded, for routine applications of QTL mapping in plant breeding programs, it is mandatory that we move away from the analysis of large segregating populations of biparental crosses and develop new QTL methods and software, which are directly applicable to the genetic materials routinely tested by plant breeders.

1.4 QTL Analysis for Pre-Breeding

Wild species are mostly used for the introgression of disease resistance genes. In general, they are hardly used to improve quantitative traits in breeding programs. The main difficulties come from, in theory, the large number of genes involved, most of them with small effects, and the high costs of recovering the genetic background of the receptor cultivar. Can QTL analysis of agronomic traits provide the necessary information to find an efficient breeding strategy using wild relatives?

To uncover and access desirable alleles from wild relatives, Tanksley and Nelson (1996) proposed a breeding strategy that seemed to maximize the efficiency of QTL mapping in tandem with MAS process itself. Bernacchi et al. (1998a,b) demonstrated how effective the advanced backcross QTL analysis strategy could be in a population of more than 300 lines derived from an initial cross between *S. hirsutum* and an elite tomato cultivar. Their results demonstrated that the wild parent carried several desirable alleles for the traits of interest, even though its own phenotype was nearly always poorer. Also reinforcing the interest on information provided by QTL analysis in advanced populations, Villalta et al. (2007) found that (1) non-correlated traits may present QTL(s) that are associated with the same marker, being responsible of correlated response, and (2) concerning the relevant genes involved, the genetic control of a trait may change in absence of significant G × E interaction. Similar results were found by Malmberg et al. (2005) for fitness-related traits using *Arabidopsis* RILs because although the total number of significant additive and epistatic QTLs was similar under the two growth conditions (field and greenhouse), the map locations were largely different. In addition to the detection of favorable QTL alleles in the wild germplasm, knowledge on favorable (and unfavorable) epistasis and their dependence on the environmental conditions should be considered.

The need for integrating information from multiple traits and environments makes fixed segregant populations important in QTL analysis. Three kinds of fixed segregant populations are generally used for this purposes: NIL (near isogenic lines, the final products of the above mentioned advanced backcross lines when wild genome fragments are covering the whole genome in separate lines), DHL (doubled haploid lines) and RIL (recombinant inbred lines). DHL and RIL mainly differ in the number of recombination cycles (making a large difference in mapping resolution) and the need to use tissue culture protocols to obtain plants (DHL) from F1 gametes.

A disadvantage of advanced backcross design is that this kind of populations is not useful for detecting epistatic QTL (Tanksley and Nelson 1996; Asins 2002) since every backcross generation greatly reduces the number of genotypic combinations because the donor genotype is being recovered. Similarly to advanced backcross lines, near-isogenic lines (NIL) are less powerful to detect epistatic QTLs and show lower localization resolution than RIL (Keurentjes et al. 2007). An experimental design that facilitates the study of epistasis using wild germplasm to increase adaptation might be important for several reasons. At the biochemical level, epistatic interactions are regulating metabolic pathways (Rowe et al. 2008). At phenotypic level, two studies in *Arabidopsis* (Malmberg et al. 2005) and rice (Mei et al. 2005)

suggest that epistatic QTL effects are more important than additive QTL for fitness traits. By contrast, studies in maize revealed that epistasis was of little or only moderate importance which following Holland (2007) might be explained in part to the relative importance of epistatic effects in autogamous versus allogamous species, and in part to differences in modeling procedures. Noteworthy, Estañ et al. (2008) found no common additive QTL but only one with epistatic effects for fruit yield when comparing two populations of *Solanum* RILs as rootstocks. In a previous study, Villalta et al. (2007), also comparing these two populations of RILs that share the female progenitor, found remarkable differences in the QTL location of the target traits, under both control and salinity conditions. These results reinforce the idea that connected populations enlarge the scope of natural QTL allelic diversity under study for agronomic traits, which is a good choice for future germplasm utilization.

Nevertheless, the RIL design present several problems. Thus, the development of a sufficiently large population of RILs is not equally easy for all autogamous species. Two populations of F_6 lines from crosses of *S. lycopersicum* var. cerasiforme as female parent with *S. pimpinellifolium* (P population) and with *S. cheesmanii* (C population) as male parents were developed by single seed descent from 300 or 400 F_2 individual plants (Monforte et al. 1997a), respectively, with no conscious selection at any generation, under greenhouse or screen house conditions. In spite of starting with 100 F_2 plants more, the number of F_6 lines was smaller in the C population (115) than in the P population (142). Main differences found between F_2 and F_6 populations were: (1) the large increment of genotypic segregation distortion from generation 2 to 6 and (2) the origin of this distortion was an excess of homozygotes (for *P* or *L* alleles) in F_2 while in the F_6, the distortion was mostly due to an excess of heterozygotes (Villalta et al. 2005). Therefore, as the number of generations of self-pollination increased, the viability and/or fertility of homozygotes (at any genomic location) decreased making allele fixation difficult. Reduction of the frequency of the wild allele and/or the maintenance of a high level of heterozygosity as found in the C population had also been observed in the development of other advanced populations of tomato interspecific crosses involving *S. cheesmanii* (Paran et al. 1995), *S. peruvianum* (Fulton et al. 1997) and *S. hirsutum* (Monforte and Tanksley 2000). Since line genotyping is usually carried out using DNA from pooled individuals, the high level of heterozygosity complicates the QTL analysis. Thus, Villalta et al. (2007) found that results on QTL detection changed if heterozygous class was considered or not. Additionally, a major problem arose when interval-mapping methodology was used allowing dominance because new QTLs appeared that were not significant under strict additivity or using methodologies based on one marker at a time. In addition to the unusual level of heterozygosis and the deviation from the completely additive QTL model, the distribution of most traits deviated from normality. Since violating the normality assumption has an impact on the distribution of the statistic used to test for a QTL, it makes also standard statistical procedures potentially inaccurate (Doerge 2002), and other, non-parametric tests for QTL detection were used (Kruglyak and Lander 1995) loosing precision in the location.

Another disadvantage of RIL population, pointed out by Keurentjes et al. (2007), is that smaller-effect QTL could be detected in the NIL population although the

localization resolution was lower. In fact, a limitation of the RIL design is that linkage disequilibrium affects much shorter genomic regions than in other segregating populations that are based on fewer recombination cycles such as F_2, DHL or even NIL. Therefore, for a similar marker density, fewer QTLs are expected to be detected in the RIL population (Villalta et al. 2007). Will higher mapping resolution also affect the estimates of QTL contribution? The cumulative recombination along the inbreed generations changes some genetic parameters of quantitative traits. It is well known (Falconer 1960) that the reduction of heterozygosis with inbreeding causes a redistribution of genetic variance in terms of genetic differentiation between lines and genetic uniformity within lines. Thus, when inbreeding is complete and the interaction of genetic variance arising from epistasis is neglected, the genetic variance in the population as a whole is doubled, and all of it appears in the between line component. Therefore, if QTLs detected at F_2 correspond to single genes (and not to blocks of genes), we should expect higher PVEs in F_7 than in F_2. Since we observed that estimated PVEs at F_7 are the lowest, at least one of the assumptions (single genes instead of blocks of genes, and lack of epistasis) must fail. Numerous significant epistatic interactions were found that varied depending on the wild species and the level of salinity (Asins et al., unpublished results) but they are very difficult to evaluate in comparison to the contribution of single additive QTL effects for the sample size used and the level of heterozygosis found. On the other hand, how often a QTL represent a block of genes? In studies with model organisms (Cohen et al. 2000; Blumenthal et al. 2002; Lercher et al. 2002) functionally interacting genes are often closely located and tightly linked on a chromosome. In tomato, Monforte et al. (2001), and Chen and Tanksley (2004) have reported QTLs likely to be composed of multiple linked genes. But, on the other hand, several single genes have already been identified as responsible of QTLs (Korstanje and Paigen 2002; Morgante and Salamini 2003; Paran and Zamir 2003; Price 2006). Another, non-excluding possibility, is that the overestimation contributions would be related not only to the presence of dominance but also to the low mapping resolution (such as in the F_2 study by Monforte et al. 1997a). If it was so, fixed segregant populations, such as DHLs and NILs that have lower mapping resolution than RILs might also induce such biased estimates. This would explain results by Keurentjes et al. (2007) who found that the total number of QTLs detected did not differ much between NIL and RIL populations but the average explained variance of a single QTL was higher in the NIL population.

1.5 An Example of QTL Analyses for Utilization of Wild Species to Increase Salt Tolerance in Tomato

Tomato is one of the most important horticultural crops. In terms of human health, tomato fruit is a major component of daily meals in many countries and constitutes an important source of minerals, vitamins, and antioxidant compounds. However the areas for tomato optimal growing conditions are becoming narrower around the world.

About 20% of irrigated agricultural land and 2% of dry land agriculture are affected by salinity (Yeo et al. 1999). Since salt tolerance, such as tolerance to any abiotic stress, means adaptation, breeding for salt tolerance should take advantage of the evolution of *Solanum* species occurred through adaptation to marginal environments. In this sense, two tomato wild species have been considered as possible donors of salt tolerance: *S. pimpinellifolium* L. (Bolarín et al. 1991; Cuartero et al. 1992; Asins et al. 1993; Foolad and Lin 1997) and *S. cheesmaniae* (L. Riley) Fosberg (Rush and Epstein 1976; Tal and Shannon 1983; Mahmoud et al. 1986; Asins et al. 1993). However, in spite of the great effort devoted to breed for salt tolerance, only a small number of cultivars, partially tolerant to salinity, have been developed (Owen et al. 1994; Al-Doss and Smith 1998; Dierig et al. 2001; Steiner and Banuelos 2003). Two major problems are encountered: the definition, or selection criteria, for "salt tolerance" and the efficient use of the wild germplasm to increase the salt tolerance of the crop.

Efforts on salt tolerance dissection using tomato experimental populations have been carried out taking into account different kinds of traits; however, in the case of crop plants, it is ultimately the yield under specific field conditions that will determine whether or not a gene or combination of genes (or QTLs) is of agronomic importance. In a previous study, we compared QTLs involved in the salt tolerance, in terms of fruit yield, of three F_2 families derived from crosses of the tomato cultivated species (*Solanum lycopersicum* L.) and two wild species *S. pimpinellifolium* L. and *S. cheesmaniae* (L. Riley) Fosberg (Monforte et al. 1997a, b). Nevertheless, breeding for salt tolerance requires breeding for a wide adaptation to different salinity levels, because salinity is an abiotic stress factor that varies in time and space. Since the possibilities to study the genetics and physiology of adaptation to different salinity levels in terms of G × E interaction were found to be very limited in F_2 populations, we developed and genetically characterized two populations of RILs derived from those F_2 populations (Villalta et al. 2005). Besides, a better understanding of the whole plant behavior under changing salinity levels was needed to improve salt tolerance efficiently. This knowledge should involve not only the trait defined as salt tolerance but also other correlated traits because correlated responses may constrain the success of breeding programmes. Salt tolerance in terms of fruit yield was then studied by QTL analysis using two RIL populations (Villalta et al. 2007) and contrary to expected, it was found that the wild allele (i.e. from the wild salt tolerant genotype) was advantageous only at one total fruit yield QTL on chromosome 10 (*tw10.1*, near the salt specific *fn10.1*). In fact, they found that the advantageous allele at all fruit weight QTLs came from the cultivated, salt sensitive, species. Moreover, when QTL controlling Na^+ and K^+ leaf accumulation, as physiological components of salt tolerance, were investigated in these populations, only two sodium QTLs (*lnc1.1* and *lnc8.1*) mapped in genomic regions where fruit yield QTLs were also located. In both cases, the profitable allele corresponded to the salt sensitive, cultivated species. Therefore, other approaches to raising tolerance to salt using wild germplasm needed to be considered.

The grafting technique has been used in agriculture since ancient time to improve horticultural crops. Nonetheless, the mechanism by which the rootstock affects the scion trait remains elusive. In relation to salinity, Estañ et al. (2005) had shown that

grafting raised fruit yield of a tomato hybrid variety under salinity. This moved us to genetically study the rootstock effect on the fruit yield of a grafted tomato variety under salinity using as rootstock the same two populations of RILs now at F_9. Since these populations at F_7, had been previously used for QTL analysis of fruit yield of the non-grafted lines allowing the comparison of genetic parameters between grafted and non-grafted plants (Estañ et al. 2008). Main results were: (1) there were rootstock lines from the two populations (up to 65% in the P population) that raised the fruit yield of the commercial hybrid under saline conditions; (2) this salt tolerance rootstock effect is a heritable trait (h^2 near 0.3), governed by at least eight QTLs; (3) most detected QTLs corresponded to the number of fruits, in agreement with the major relevance of this component among rootstock effects on fruit yield; (4) in general, QTL gene effects were medium-sized, with contributions from 8.5% up to 15.9% at most, and the advantageous allele came from the wild, salt tolerant, species; (5) only two fruit yield QTLs on chromosomes P9 and C11 might correspond to fruit yield QTLs of the non-grafted lines indicating their root system dependence; and (6) no common QTL between population was found but a fruit yield QTL on chromosome 3 was acting epistatically in both populations.

The fact that a certain proportion of lines increased the fruit yield of the grafted hybrid variety under salinity was promising but since salinity is variable in time and space it was important to test the lack of negative effects on yield under non saline irrigation. This experiment was carried out with 50 lines from each population and two controls: the non-grafted variety (Boludo, Bol) and the self-grafted variety (Bol/Bol). As shown in Fig. 1.1, most lines are similar than controls in absence of salinity. Therefore, an easy, efficient and profitable utilization of wild germplasm can be carried out through the improvement of rootstocks that confer salt tolerance in terms of fruit yield to the grafted variety instead of introgression their beneficial QTL alleles into the genome of the cultivated tomato.

1.6 Concluding Remarks

The new challenges of plant breeding urge to integrate the latest innovations in biology and genetics to enhance crop improvement. QTL analysis fills the gap between "omics" and the field. Since most cloned QTLs are elements involved in gene regulation, the candidate approach, comparing the position of plant phenotype QTLs, gene expression QTLs (eQTLs) and network eQTLs may increase our chances to clone QTLs of agronomic traits. This vertical integration of information needs the availability of fixed segregant populations. Recombinant inbred line populations constitute a valuable experimental design for this purpose because (1) their mapping resolution (important to distinguish between linkage from pleiotropy) and (2) they facilitate the study of epistasis. Epistatic QTLs are frequently missed when using other experimental designs.

QTL meta-analysis is a horizontal way to integrate information for the same trait from different experiments, populations (better if they are connected through com-

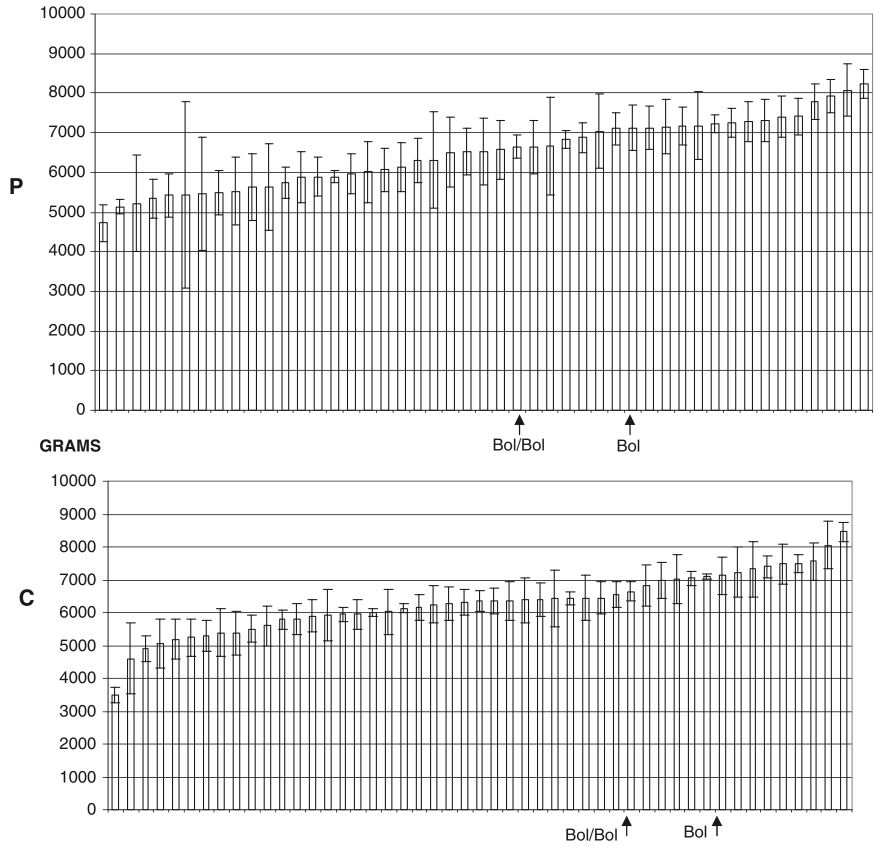


Fig. 1.1 Ordered means with standard errors for total fruit weight (in grams) of Boludo (Bol) using some lines of the P and the C populations as rootstock under optimal growing conditions in comparison to the controls, non-grafted and self-grafted Boludo plants (Bol and Bol/Bol). Arrows indicate the relative position of these controls.

mon parents) and closely related species. Therefore, if statistical limitations are overcome, it will benefit comparative genomics and it will be a valuable tool to complement information obtained from vertical integration. The integration of QTL analyses of salt tolerance in tomato using wild relatives as donors of salt tolerance has been useful to judge between breeding strategies. Their utilization in rootstock breeding programs seems the most efficient one because most "good-rootstock" QTL alleles come from the wild species. Association mapping methods can be used to validate QTLs by testing the effects of linked markers or candidate genes within a large set of genotypes such as germplasm accessions or cultivated accessions. These or other methods of cross validation of QTL mapping results (particularly, allele effects) are absolutely necessary to ensure the success of marker-assisted selection.

Instead of inferring MAS schemes in breeding populations from QTL results in experimental (bi-parental) populations, the development of QTL mapping methods

for multiparental populations is envisaged as a better alternative because it would allow to link QTL analysis and MAS in tandem through generations in the breeding program itself.

Acknowledgements This work was supported in part by a grant from MCI (RTA04-075-C2 and RTA2006-0009-00-00). Authors thank Dr. Alfredo Miguel from IVIA and Dr. Carlos Biaxial from Foundation "Ruralcaja" of Valencia for the grafting, culture and evaluation of the *Solanum* RIL populations as rootstock under optimal growing conditions.

References

Al-Doss AA, Smith SE (1998) Registration of AZ-97MEC and AZ-97MEC-ST very nondormant alfalfa germplasm pools with increased shoot weight and differential response to saline irrigation. Crop Sci 38:568–568

Asins MJ (2002) Present and future of QTL analysis in plant breeding. Plant Breed 121:281–291

Asins MJ, Breto MP, Carbonell EA (1993) Salt tolerance in *Lycopersicon* species. II. Genetic effects and search of associated traits. Theor Appl Genet 86:769–774

Ballini E, Morel JB, Droc G, Price A, Courtois B, Notteghem JL, Tharreau D (2008) A genome-wide meta-analysis of rice blast resistance genes and quantitative trait loci provides new insights into partial and complete resistance. Mol Plant Microbe Interact 21:859–868

Beavis WD (1994) The power and the deceit of QTL experiments: Lessons from comparative studies. 49th Annual Corn and Sorghum Industry Research Conference. American Seed Trade Association, Washington, DC, pp 250–266

Beavis WD, Schilkie FD, Baxter SM (2007) Translational bioinformatics: at the interface of genomics and quantitative genetics. Crop Sci 47:S32–S43

Benjamini Y, Yekutieli D (2005) Quantitative trait loci analysis using the false discovery rate. Genetics 171:783–790

Bernacchi D, Beck-Bunn T, Emmatty D, Eshed Y, Inai S, Lopez J, Petiard V, Sayama H, Uhlig I, Zamir D, Tanksley S (1998a) Advanced backcross QTL analysis of tomato. II. Evaluation of near-isogenic lines carrying single-donor introgressions for desirable wild QTL-alleles derived from *Lycopersicon hirsutum* and *L. pimpinellifolium*. Theor Appl Genet 97:170–180

Bernacchi D, Beck-Bunn T, Eshed Y, Lopez J, Petiard V, Uhlig I, Zamir D, Tanksley S (1998b) Advanced backcross QTL analysis of tomato. I. Identification of QTLs for traits of agronomic importance from *Lycopersicon hirsutum*. Theor Appl Genet 97:381–397

Bernardo R, Charcosset A (2006) Usefulness of gene information in marker-assisted recurrent selection: a simulation appraisal. Crop Sci 46:614–621

Bernet GP, Gorris MT, Carbonell EA, Cambra M, Asins MJ (2008) *Citrus tristeza virus* resistance in a core collection of sour orange based on a diversity study of three germplasm collections using QTL-linked markers. Plant Breed 127:398–406

Blanc G, Charcosset A, Mangin B, Gallais A, Moreau L (2006) Connected populations for detecting quantitative trait loci and testing for epistasis: an application in maize. Theor Appl Genet 113:206–224

Blumenthal T, Evans D, Link CD, Guffanti A, Lawson D, Thierry-Mieg J, Thierry-Mieg D, Chiu WL, Duke K, Kiraly M, Kim SK (2002) A global analysis of *Caenorhabditis elegans* operons. Nature 417:851–854

Bolarin MC, Fernandez FG, Cruz V, Cuartero J (1991) Salinity tolerance in 4 wild tomato species using vegetative yield salinity response curves. J Am Soc Hortic Sci 116:286–290

Brummer EC (2004) Applying genomics to alfalfa breeding programs. Crop Sci 44:1904–1907.

Campbell BT, Baenzinger PS, Gill KS, Eskridge KM, Budak H, Erayman M, Yen Y (2003) Identification of QTLs and environmental interactions associated with agronomic traits on chromosome 3A of wheat. Crop Science 43:1493–1505

Causse M, Chaïb J, Lecomte L, Buret M, Hospital F (2007) Both additivity and epistasis control the genetic variation for fruit quality traits in tomato. Theor Appl Genet 115:429–442

Chardon F, Virlon B, Moreau L, Falque M, Joets J, Decousset L, Murigneux A (2004) Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome

Chen KY, Tanksley SD (2004) High resolution mapping and functional analysis of se2.1: A major stigma exsertion quantitative trait locus associated with the evolution from allogamy to autogamy in the genus *Lycopersicon*. Genetics 168:1569–1579

Cohen BA, Mitra RD, Hughes JD, Church GM (2000) A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nat Genet 26:183–186

Collard BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Phil Trns R Soc B 363:557–572

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basis concepts. Euphytica 142:169–196

Cuartero J, Yeo AR, Flowers TJ (1992) Selection of donors for salt-tolerance in tomato using physiological traits. New Phytol 121:63–69

Dierig DA, Shannon MC, Grieve CM (2001) Registration of WCL-SL1 salt tolerant *Lesquerella fendleri* germplasm. Crop Sci 41:604–605

Doerge RW (2002) Mapping and analysis of quantitative trait loci in experimental populations. Nat Rev/Genet 3:43–52

Doerge RW, Rebai A (1996) Significance thresholds for QTL mapping tests. Heredity 76:459–464

Estañ MT, Martinez-Rodriguez MM, Pérez-Alfocea F, Flowers TJ, Bolarín MC (2005) Grafting raises the salt tolerance of tomato through limiting the transport of sodium and chloride to the shoot. J Exp Bot 56:703–712

Estañ MT, Villalta I, Bolarín MC, Carbonell EA, Asins MJ (2008) Identification of fruit yield loci controlling the salt tolerance conferred by solanum rootstocks. Theor Appl Genet DOI: 10.1007/s00122–008–0900–6

Falconer DS (1960) Introduction to Quantitative Genetics, R. Mac Lehose and com., Glasgow, pp 264–275

Foolad MR, Lin GY (1997) Genetic potential for salt tolerance during germination in *Lycopersicon* species. Hortscience 32:296–300

Fulton TM, Nelson JC, Tanksley SD (1997) Introgression and DNA marker analysis of *Lycopersicon peruvianum*, a wild relative of the cultivated tomato, into *Lycopersicon esculentum*, followed through three successive backcross generations. Theor Appl Genet 95:895–902

Göring HHH, Terwillinger JD, Blangero J (2001) Large up-ward bias in estimation of locus-specific effects from genomewide scans. Am J Hum Genet 69:1357–1369

Grandillo S, Ku HM, Tanksley SD (1999) Identifying the loci responsible for natural variation in fruit size and shape in tomato. Theor Appl Genet 99:978–987

Guimarães EP, Kueneman E (2006) Assesment of national plant breeding and biotechnology capacity worldwide. HortScience 41:50–52

Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol Biol 57:461–485

Gupta PK, Rustgi S, Mir RR (2008) Array-based high-throughput DNA markers for crop improvement. Heredity 101:5–18

Hansen B, Halkier BA, Kliebenstein D (2008) Identifying the molecular basis of QTLs: eQTLs add a new dimension. Trends Plant Sci 13:72–77

Hazen SP, Wu YJ, Kreps JA (2003) Gene expression profiling of plant responses to abiotic stress. Func Integr Genomics 3:105–111

Holland JB (2007) Genetic architecture of complex traits in plants. Curr Opin Plant Biol 10:156–161

Kearsey MJ, Farquhar GL (1998) QTL analysis in plants; where are we now? Heredity 80:137–142

Kearsy MJ, Pooni HS, Syed NH (2003) Genetics of quantitative traits in *Arabidopsis thaliana*. Heredity 91:456–464

Keurentjes JJB, Bentsink L, Alonso-Blanco C, Hanhart CJ, Blankestijn-de Vries H, Effgen S, Vreugdenhil D, Koornneef M (2007) Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. Genetics 175:891–905

Kliebenstein DJ, West MAL, van Leeuwen H, Loudet O, Doerge RW, St Clair DA (2006) Identification of QTLs controlling gene expression networks defined a priori. BMC Bioinformatics 7

Knight J (2003) Crop improvement: a dying breed. Nature 421:568–570

Korstanje R, Paigen B (2002) From QTL to gene: the harvest begins. Nat Genet 31:235–236

Kruglyak L, Lander ES (1995) A non-parametric approach for mapping quantitative trait loci. Genetics 139:1421–1428

Lan H, Chen M, Flowers JB, Yandell BS, Stapleton DS, Mata CM, Mui ET-K, Flowers MT, Schueler KL, Manly KF, Williams RW, Kendziorski C, Attie AD (2006) Combined Expression Trait Correlations and Expression Quantitative Trait Locus Mapping. PLoS Genet. 2(1):e6. doi: 10.1371/journal.pgen.0020006

Langridge P, Lagudah P, Holton T, Appels R, Sharp P, Chalmers K (2001) Trends in genetic and genome analyses in wheat: a review. Aust J Agric Res 52:1043–1077

Lecomte L, Duffé P, Buret M, Servin B, Hospital F, Causse M (2004) Marker-assisted introgression of five QTLs controlling fruit quality traits into three tomato lines revealed interactions between QTLs and genetic backgrounds. Theor Appl Genet 109:658–668

Lercher MJ, Urrutia AO, Hurst LD (2002) Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nat Genet 31:80–183

Lisec J, Meyer RC, Steinfath M, Redestig H, Becher M, Witucka-Wall H, Fiehn O, T÷rjqk O, Selbig J, Altmann T, Willmitzer L (2008) Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations. Plant J 53: 960–972

Lu C, Shen L, Tan Y, Xu Y, He P, Chen Y, Zhu L (1996) Comparative mapping of QTL for agronomic traits of rice across environments using a double haploid population. Theor Appl Genet 93:1211–1217

Luo ZW, Potokina E, Druka A, Wise R, Waugh R, Kearsey MJ (2007) SFP genotyping from Affymetrix arrays is robust but largely detects cis-acting expression regulators. Genetics 176:789–800

Mahmoud MH, Jones RA, El-Betagy AS (1986) Comparative responses to high salinity between salt sensitive and salt tolerant genotypes of tomato. Acta Hort 190:533–543

Malmberg RL, Held S, Waits A, Mauricio R (2005) Epistasis for fitness-related quantitative traits in *Arabidopsis thaliana* grown in the field and in the greenhouse. Genetics 171:2013–2027

Mei HW, Li ZK, Shu QY, Guo LB, Wang YP, Yu XQ, Ying CS, Luo LJ (2005) Gene actions of QTLs affecting several agronomic traits resolved in a recombinant inbred rice population and two backcross populations. Theor Appl Genet 110:649–659

Melchinger AE, Utz HF, Schön CC (2004) QTL analyses of complex traits with cross validation, bootstrapping and other biometric methods. Euphytica 137:1–11

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Monforte A, Tanksley SD (2000) Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: a tool for gene mapping and gene discovery. Genome 43:803–813

Monforte A, Asíns MJ, Carbonell, EA (1997a) Salt tolerance in *Lycopersicon* species. V. May genetic variability at quantitative trait loci affect their analysis? Theor Appl Genet 95:284–293

Monforte A, Asins MJ, Carbonell EA (1997b) Salt tolerance in *Lycopersicon* species. VI. Genotype by salinity interaction in quantitative trait loci detection. Constitutive and response QTLs. Theor Appl Genet 95:706–713

Monforte AJ, Friedman E, Zamir D, Tanksley SD (2001) Comparison of a set of allelic QTL-NILs for chromosome 4 of tomato: Deductions about natural variation and implications for germplasm utilization. Theor Appl Genet 102:572–590

Moose SP, Mumm RH (2008) Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiol 147:969–977

Moreau L, Charcosset A, Gallais A (2004) Use of trial clustering to study QTL × environment effects for grain yield and related traits in maize. Theor Appl Genet 110:92–105

Morgante M, Salamini F (2003) From plant genomics to breeding practice. Curr Opin Biotechnol 14:214–219

National Research Council (2008) Achievements of the national plant genome initiative and new horizons in plant biology. National Academies Press, Washington, DC

Openshaw S, Frascaroli E (1997) QTL detection and marker-assisted selection for complex traits in maize. Annu Corn Sorghum Res Conf Proc 52:44–53

Owen PA, Nickell CD, Noel GR, Thomas DJ, Frey K (1994) Registration of saline soybean. Crop Sci 34:1689–1689

Pajerowska KM, Parker JE, Gebbardt C (2005) Potato homologues of Arabidopsis thaliana genes functional in defense signaling-identification, genetic mapping and molecular cloning. Mol Plant Microbe Interact 18:1107–1119

Paran I, Zamir D (2003) Quantitative traits in plants: beyond the QTL. Trends Genet 6:303–306

Paran I, Goldman IL, Tanksley SD, Zamir D (1995) Recombinant inbred lines for genetic mapping in tomato. Theor Appl Genet 90:542–548

Patterson AH, Damon S, Hewitt JD, Zamir D, Rabinowitch HD, Lincoln SE, Lander ES, Tanksley SD (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations and environments. Genetics 127:181–197

Podlich DW, Winkler CR, Cooper M (2004) Mapping as you go: an effective approach for marker-assisted selection of complex traits. Crop Sci 44:1560–1571

Price A (2006) Believe it or not, QTLs are accurate! Trends Plant Sci 11:213–216

Ribaut JM, Betran J (1999) Single large-scale marker-assisted selection (SLS-MAS). Mol Breed 5:531–541

Rowe HC, Hansen BG, Halkier BA, Kliebenstein DJ (2008) Biochemical networks and epistasis shape the *Arabidopsis thaliana* metabolome. Plant Cell 20:1199–1216

Rush DW, Epstein E (1976) Genotypic responses to salinity.Differences between salt-sensitive and salt-tolerant genotypes of tomato. Plant Phys 57:162–166

Salvi S, Tuberosa R (2005) To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci 10:297–304

Sari-Gorla M, Calinski T, Kaczmarek Z, Krajewski P (1997) Detection of QTL × environment interaction in maize by a least squares interval mapping method. Heredity 78:146–157

Sen S, Churchill GA (2001) A statistical framework for quantitative trait mapping. Genetics 159:371–387

Slade AJ, Fuerstenberg SI, Loeffier D, Steine MN, Facciotti D (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nat Biotechnol 23:75–81

Sölkner J, Grausgruber H, Okeyo AM, Ruckenbauer P, Wurzinger M (2008) Breeding objectives and the relative importance of traits in plant and animal breeding: a comparative review. Euphytica 161:273–282

Steiner JJ, Banuelos GS (2003) Registration of ARS NLT-SALT and ARS-NLT-SALT/B saline tolerant narrow-leaf trefoil germplasm. Crop Sci 43:1888–1889

Stuber CW, Lincoln SE, Wolf DW, Helentjaris T, Lander ES (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823–839

Sun W, Yu TW, Li KC (2007) Detection of eQTL modules mediated by activity levels of transcription factors. Bioinformatics 23:2290–2297

Tal M, Shannon MC (1983) Effects of dehydration and high-temperature on the stability of leaf membranes of *Lycopersicon esculentum*, *L. cheesmanii*, *Lycopersicon peruvianum* and *Solanum pennellii*. Z Pflanzenphysiol 112:411–416

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Gene. 92:191–203

Vergne E, Ballini E, Droc G, Tharreau D, Nottqghem JL, Morel JB (2008) ARCHIPELAGO: a dedicated resource for exploiting past, present, and future genomic data on disease resistance regulation in rice. Mol Plant Microbe Interact 21:869–878

Veyrieras JB, Camus-Kulandaivelu L, Gouesnard B, Manicacci D, Charcosset A (2007) Bridging genomics and genetic diversity: linkage disequilibrium structure and association mapping in maize and other cereals. Crop Sci 47:560–571

Villalta I, Reina-Sanchez A, Cuartero J, Carbonell EA, Asins MJ (2005) Comparative microsatellite linkage analysis and genetic structure of two populations of F-6 lines derived from *Lycopersicon pimpinellifolium* and *L. cheesmanii*. Theor Appl Genet 110:881–894

Villalta I, Bernet GP, Carbonell EA, Asins MJ (2007) Comparative QTL analysis of salinity tolerance in terms of fruit yield using two solanum populations of F7 lines. Theor Appl Genet 114:1001–1017

Villalta I, Reina-Sánchez A, Bolarín MC, Cuartero J, Belver A, Venema K, Carbonell EA, Asins MJ (2008) Genetic analysis of Na^{+} and K^{+} concentrations in leaf and stem as physiological components of salt tolerance in tomato. Theor Appl Genet 116:869–880

Vuylsteke M, van den Daele H, Vercauteren A, Zabeau M, Kuiper M (2006) Genetic disection of transcriptional regulation by cDNA-AFLP. Plant J 45:429–446

Wayne ML, McIntyre L (2002) Combining mapping and arraying: an approach to candidate gene identification. PNAS USA 99:14903–14906

West MAL, van Leeuwen H, Kozik A, Kliebenstein DJ, Doerge RW, St. Clair DA, Michelmore RW (2006) High-density haplotyping with microarray-based expression and single feature polymorphism markers in *Arabidopsis*. Genome Res 16:787–795

Wisser RJ, Balint-Kurti PJ, Nelson RJ (2006) The genetic architecture of disease resistance in maize: a synthesis of published studies. Phytopathology 96:120–129

Xu Y, Crouch JH (2008) Marker-assisted selection in plant breeding: from publications to practice. Crop Sci 48:391–407

Yan J, Zhu J, He C, Benmossa M, Wu P (1999) Molecular marker-assisted dissection of genotype x environment interaction for plant type traits in rice (*Oryza sativa* L.) Crop Sci 39:538–544

Yeo AR, Flowers SA, Rao G, Welfare K, Seanayake N, Flowers TJ (1999) Silicon reduces sodium uptake in rice (*Oryza sativa* L.) in saline conditions and this is accounted for by a reduction in the transpirational flow. Plant Cell Env 22:559–565

Young ND (1999) A cautiously optimistic vision for marker-assisted breeding. Mol Breed 5:505–510

Zou W, Zeng Z-B (2008) Statistical methods for mapping multiple QTL. Int J Plant Genomics. doi:10.1155/2008/286561

Chapter 2
Comparative Genomics in Crop Plants

Mehboob-ur-Rahman and Andrew H. Paterson

Abstract Angiosperms evolved from a common ancestral genome that incurred repeated duplications and many mutations in succeeding generations to result in the evolution of an array of plant species, setting the stage for the application of comparative genomic approaches to its descendant modern angiosperms. Presently, much genomic information including complete genome sequences from model crops is available, with partial genomic information for many other plant taxa shedding light on genome structure and gene repertoire. Comparative approaches permit inferences to be made regarding evolutionary consequences including rates of evolution of particular genes or families, differential gene loss or retention following duplications, and chromosomal rearrangements, collectively contributing to taxonomic, morphological, and physiological variation. The expanding genomic information for angiosperms may soon permit us to deduce ancestral genome karyotypes for each plant family, and perhaps even for the common ancestor of all angiosperms. Future availability of additional sequenced genomes coupled with complementary bioinformatics tools may help to redraw the plant phylogeny, identify the ancestral gene set for angiosperms and clarify the subsequent evolutionary history of these genes, and provide new insight into the causes and consequences of fluctuating genome size. Better understanding of relationships among different angiosperm genomes and their constituent genes will expedite goals ranging from isolation of genes and determination of their functions, to identifying DNA markers useful for marker-assisted breeding.

2.1 Introduction

Variation in different characters of living organisms is the link between DNA and phenotype, providing for selection to act to improve fitness in a particular environment (whether natural or agricultural), and lying at the heart of both the Darwinian

Mehboob-ur-Rahman
Plant Genomics and Mol Breeding Labs, National Institute for Biotechnology and Genetic Engineering (NIBGE), Faisalabad, Pakistan

A.H. Paterson (✉)
Plant Genome Mapping Laboratory, University of Georgia, Athens, Georgia, 30602, USA
e-mail: paterson@dogwood.botany.uga.edu

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_2,

evolutionary principles and the Mendelian genetic principles. Comparison of traits between two organisms, especially within a plant species, started with the proposition of Vavilov's law of homologous series in variation. In the last two decades, especially with the availability of genetic maps and genome sequence information, the concept was broadened to include comparisons of genomes structure and function that are now referred to as comparative genomics.

Key to plant comparative genomic studies is that the locations of some genes has remained stable across long evolutionary times, providing a framework useful for inferring correspondence among even distantly-related genomes, and among genes that have diverged beyond recognition or been relocated (Tang et al. 2008a). However, the degree to which the genes remain on corresponding chromosomes (synteny) and in corresponding orders (collinearity) over time differs markedly among taxa (Coghlan et al. 2005), with angiosperms incurring differential retention of genes and or gene families following repeated genome duplications and mobility of DNA sequences (Bowers et al. 2005; Tang et al. 2008a). Determining the extent of collinearity between model plant species and related crops is valuable for applying genomic information to crop improvement.

Much progress has been made in linking plant genomes through comparative genetic/physical maps, especially for species belonging to the same family. For example, close relationships have been revealed among the genomes of many members of the grass family, among the Solanaceae (nightshade) crops, among the Brassicaceae (cole) crops, and among several legume crops (Paterson et al. 2000; Wu et al. 2006), facilitating reconstruction of ancestral genomes (Blanchette et al. 2004), phylogenetic studies (Rokas et al. 2003), deciphering of patterns of natural selection on coding regions (Bustamante et al. 2005), transferring predictions of common gene function between the species (Eisen 1998; Doganlar et al. 2002b) etc., and other applications. Integration of sequencing and detailed functional analysis of sexually incompatible species is advantageous to both the geneticist and breeder for better understanding and improving many understudied crops (Gale and Devos 1998) which was practically demonstrated on 22 genotypes representing eight different genera of the Poaceae family (Feltus et al. 2006). Comparative alignment of chromosomes across the species has shown that quantitative trait loci (QTLs) and major genes often occur in corresponding locations (Lin et al. 1995; Paterson et al. 1995; Pereira and Lee 1995), perhaps helping in isolating candidate genes for corresponding QTLs in different species (Lagercrantz et al. 1996).

2.2 Transition in Concept from Wet to Dry Lab

Many foundations of biostatistics are rooted in agricultural sciences, particularly agricultural genetics. Exponentially-growing quantities of DNA sequence, gene expression, nucleotide variation and other genetic and genomic data have compelled researchers to add new dimensions to systems for analyzing the data. Indeed, the availability of enormous data sets via the internet, and their comparison in novel

ways, has led to the birth of a new discipline, "bioinformatics", transitioning away from empirical (wet-lab) studies of plant biology to more emphasis on "dry lab" studies using computer science, engineering and mathematical methodologies to manage, visualize, and analyze voluminous data to discover new patterns and build hypotheses and models (Rhee 2005). In angiosperms, bioinformatics has increasingly been used for understanding the evolution of genomes, their structure and function, phylogenetic studies and predicting gene function (Paterson and Bennetzen 2001; Taher et al. 2004).

A number of dedicated genome databases provide repositories for expanding genetic information and also allow researchers to choose their preferred organism for comparisons and annotations. A few of the more prominent examples are described hereafter.

Gramene is a clade-oriented database for grasses (Beavis et al. 2005), using the rice genome as a foundation (http://www.gramene.org) (Ware et al. 2002; Jaiswal et al. 2006; Liang et al. 2008). Gramene's core data types include genome assembly and annotations, DNA/mRNA sequences, genetic and physical maps/markers, genes, QTLs, proteins, ontologies, literature and comparative mappings. The website has received much attention to make it user-friendly and also is regularly equipped with new features including rice pathways for functional annotation of rice genes; genetic diversity data from rice, maize and wheat to show genetic variations among germplasm; large-scale genome comparisons among *Oryza sativa* and its wild relatives or other taxa such as maize and sorghum for evolutionary studies; and the creation of orthologous gene sets and phylogenetic trees among several reference crop and animal species (Liang et al. 2008).

Motivated in part by the completion of the rice genome sequence, several additional rice-related databases are also available. "The Institute for Genomic Research (TIGR) Rice Genome Annotation resource" (http://rice.tigr.org) contains improved TE detection systems and gene annotation through incorporation of multiple transcript and proteomic expression data sets. Structural and functional annotations are viewable through a genome browser. Enhanced data access is available through web interfaces, FTP downloads and a Data Extractor tool developed in order to support discrete dataset downloads (Ouyang et al. 2007). The Rice Genome Automated Annotation System (RiceGAAS: http://ricegaas.dna.affrc.go.jp) is also an annotation and database tool for rice genome sequences ranging from 10 kb to 1 Mb submitted to GenBank (Sakata et al. 2002). Manually curated annotation of the Nipponbare genome sequence can be accessed through the Rice Annotation Database (RAD: http://rad.dna.affrc.go.jp) (Ito et al. 2005) and the Integrated Rice Genome Explorer (INE: http://rgp.dna.affrc.go.jp/giot/INE.html) (Sakata et al. 2000). Furthermore, INE integrates the genome sequence information with the genetic map, physical map, and transcript map of rice. Finally, Oryzabase (http://www.shigen.nig.ac.jp/rice/oryzabase/top/top.jsp) provides comprehensive information about rice development and anatomy, rice mutants, and genetic resources especially for wild rice varieties. Several genetic, physical, and expression maps with full genome and cDNA sequences have also been corroborated with biological data for understanding the life cycle of rice, the relationship between phenotype

and gene function, and rice genetic diversity. Moreover, Oryzabase publishes the Rice Genetics Newsletter (Kurata and Yamazaki 2006).

There are dedicated websites and databases available for maize. Panzea (http://www.panzea.org), the public web site, encompasses 'Molecular and Functional Diversity in the Maize Genome'. The most significant data content expansion occurred for single nucleotide polymorphisms (SNPs), sequencing, isozyme and phenotypic data types. Furthermore, making new software available, improvement in the coding system and addition of sections for educational purposes are attractive features of the website (Canaran et al. 2008). Similarly, another database "maizeGDB" contains information on maize genetics and genomics including maps, gene product information, loci and their various alleles, phenotypes (both naturally occurring and as a result of directed mutagenesis), stocks, sequences, molecular markers, references and contact information for maize researchers worldwide (Lawrence et al. 2007). It can be accessed online at http://www.maizegdb.org. Some other useful databases are http://maize.agron.iastate.edu and http://moulon.moulon.inra.fr/imgd that can be accessed for QTL studies.

Sorghum genomic and phenotypic data can be accessed through various online resources. The Comparative Saccharinae Genome Resource website http://cggc.agtec.uga.edu/ (can also be assessed through its original link http://cggc.agtec.uga.edu) focuses on comparative and evolutionary genomics of the Saccharinae (sorghum, sugarcane and their relatives) (Kresovich et al. 2005; Paterson et al. 2005). Sorghum ESTs, genetic/physical map, and polymorphism data can be accessed via web interfaces and bulk downloads (Paterson et al. 2005). Complementary web-based resources, http://fungen.botanyuga.edu focus on functional genomics of the transcriptome, and http://sorgblast2.tamu.edu deals with the genomics of abiotic stress responses of sorghum. Finally, online resources from SUCEST (http://sucest.lbi.dcc.unicamp.br/en/), an exclusive EST project in closely related sugarcane (Vettore et al. 2003), are also often of value for sorghum genomics.

The Solanaceae comprises of many important model plant species such as tomato, tobacco, potato, eggplant and pepper (Fernie and Willmitzer 2001; Pedley and Martin 2003; Giovannoni 2004; Tanksley 2004). For handling the rapidly evolving genomic data, the SOL Genomics site (SGN; http://sgn.cornell.edu) is dedicated to genetic maps and marker data, handling a large EST collection with computationally derived unigene sets, cataloging and publishing phenotypic information, and providing associated tools pertaining to QTLs (Mueller et al. 2005, 2008). The SOL homepage provides links to related sites of interest such as Tomato Expression Database (TED; Fei et al. 2004) and Tomato Genomics Resource Center (TGRC) at the University of California, Davis (http://tgrc.ucdavis.edu) (Mueller et al. 2005). Another website SGN is upgraded and provides a comparative viewer for mapping data, including genetic, physical and cytological maps. It can also be installed and adapted for other websites. Moreover, the viewer allows users to upload their own maps and compare them to other maps in the system (Mueller et al. 2008).

CottonDB (www.cottondb.org) link can be used to access information on over 355,000 gene, ESTs, and contig sequences; genetic and physical map data; 8,000 DNA primers; and 9,000 germplasm accessions. It facilitates researchers for

using CMap viewer (developed by Gramene) for conducting comparative genomic analyses (Yu et al. 2008). Another useful database, the Cotton Diversity Database (http://cotton.agtec.uga.edu) provides management tools for handling phenotypic and genotypic data for drawing inferences related to phylogenetic, genetic, and comparative genomic study. The database has a capacity to integrate the queries with comparative physical, expression profiling and BAC resources (Gingle et al. 2006).

Other useful links pertaining to structural genomic studies are 'CMD (http://www.mainlab.clemson.edu/cmd/AboutUs.shtml) provide information about the publicly available cotton microsatellites' and 'TropGENE-DB (http://tropgenedb.cirad.fr/en/cotton.html) pertains to a subset of published mapping data'. The websites dealing with functional genomic studies are "cotton functional genomics" (http://cottongenomecenter.ucdavis.edu/) and cotton fiber genomics (http://www.cottongenomics.org/). There are several other useful websites (http://cottonevolution.info/microarray; http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=cotton; http://www.genome.arizona.edu/ and www.plantgenome.uga.edu) which deal with dissemination of cotton genome resources to the cotton research community. Finally, the Cotton Portal (http://gossypium.info/) provides a convenient point of entry into many cotton genomic resources.

The Soybean Genome Database (SoyGD; http://soybeangenome.siu.edu) provides information on integrated soybean physical maps, bacterial artificial chromosome (BAC) fingerprints, and genetic maps associated with genomic data. The diploid, tetraploid, octoploid, and homologous regions were highlighted in relation to an integrated genetic and physical map. For physical mapping data the most advanced build contains 2,854 contigs that encompass 1.05 Gb and 404 high-quality DNA markers anchored to 742 contigs (Shultz et al. 2006). Another useful database is Soybase (http://soybase.org/) which provides information on genetic, phenotypic, and other information about soybean. It is equipped with CMap to provide map visualizations.

The PlantTribes database (http://fgp.huck.psu.edu/tribe.html) contains information on global classification of genes derived from all five sequenced plant genomes (*Arabidopsis thaliana*, *Carica papaya*, *Medicago truncatula*, *Populus trichocarpa* and *Oryza sativa*). The important feature of this database is that a graph-based clustering algorithm MCL (Enright et al. 2002) was used to classify all protein-coding genes of these plant species into putative gene families (called tribes) at three different clustering stringencies (Wall et al. 2008). The database allows one to explore the classification of genes, to place query sequences within the classification, and to download results for further study. The database also contains unigene sets derived from more than 200 species from the TIGR Plant Transcript Assemblies (Childs et al. 2007), which can be instrumental in comparative plant genomics.

PlantGDB (http://www.plantgdb.org/) contains information of sequenced data for green plants (Viridiplantae) and provides annotated transcript assemblies for 100 plant species. Other useful features can be traced from the link (Duvick et al. 2008). For many other plant species including legumes, cole (Brassicaceae) crops

and others, additional dedicated databases are accessible from links such as http://www.geocities.com/bioinformaticsweb/speciesspecificdatabases.htm.

2.3 Dimensions of Comparative Genomics

2.3.1 Comparative Genome Organization

Genomes of flowering plants are complex, containing varying degrees of repetitive elements and genes assembled into different sets of chromosomes (Bennett and Leitch 2003; Bennetzen et al. 2005). For example, the *Arabidopsis* genome size is 140 Mb assembled into five chromosomes, versus 4,900 Mb arranged in seven chromosomes for diploid wheat. Genes are not uniformly distributed across the genomes, either large or small, being generally rich in the terminal regions of chromosomes and generally poor but not absent in the centromeric regions (Nagaki et al. 2004).

In the well-studied grasses, the size of gene-rich regions in different genomes remains similar (Feuillet and Keller 1999). Detailed comparison of the rice sequence to a genetically-anchored sorghum physical map showed cytological-distinct euchromatin versus heterochromatin to correlate closely with gene versus repeat abundance, patterns of rice-sorghum synteny, and the frequency of recombination (Bowers et al. 2005). Synteny is highest and retroelement abundance lowest in distal portions of the chromosomes, where recombination has generally been greatest per unit DNA. Preferential preservation of microsynteny in recombinogenic regions suggests that gene rearrangement is generally deleterious, an intuitive hypothesis that has previously lacked empirical support. 'Muller's ratchet' (Muller 1932) predicts that deleterious mutations may accumulate in, and contribute to degeneration of, non-recombinogenic regions, classical examples being mammalian Y chromosomes or incipient plant sex chromosomes (Liu et al. 2004). This is consistent with a strong negative correlation between repetitive DNA content and recombinational length of rice chromosomes, and with the much greater abundance of repetitive DNA than genes in the pericentromeric regions (Bowers et al. 2005). Accelerated gene loss in recombination-poor regions of wheat (Akhunov et al. 2003), and a propensity for small insertions in centromeric regions of mammalian genomes (Bailey et al. 2001), lend further support to this idea. Gene evolution can be particularly rapid in telomeric regions of wheat chromosomes (See et al. 2006) that yield new genes (Mefford and Trask 2002).

Genome size and gene number are not proportionate in the angiosperms. For example, the genome of papaya (372 Mb) is approximately three times larger than that of than *Arabidopsis* (125 Mb) but contains fewer genes (Ming et al. 2008). The *Arabidopsis* genome has incurred two whole-genome duplications that are absent from papaya – while most duplicated genes following each of these events were lost, enough persisted to account for its higher gene count. Comparison of the genome sequences of *Arabidopsis*, rice, poplar, grapevine and papaya elucidates 13,311 genes that appear to comprise a common set among these angiosperms (Ming et al. 2008).

2.3.2 Gene Prediction

One early goal after getting a draft or finished genome sequence is to explore the sequence for its particular set of genes and transcribed genomic sequences (Windsor and Mitchell-Olds 2006). Gene sets can be generated either entirely computationally or by a combination of computational and manual annotation. While the former is useful in generating preliminary gene sets, manual annotation, though laborious to perform, is still necessary for ensuring accuracy and completeness of gene sets. In this regard, different gene prediction programs, including EAnnot (Electronic Annotation) offers an efficient way to generate an automated gene set. EAnnot builds gene models based on mRNA, EST, and protein alignments to genomic sequence, attaches supporting evidence to the corresponding genes, identifies pseudogenes, and locates poly(A) sites and signals (Ding et al. 2004).

Translation of genetic information from model organisms is extremely useful in annotating conserved regions of other genomes. In angiosperms, sequencing of the *Arabidopsis* (*Arabidopsis thaliana*), rice, papaya, sorghum, poplar, and grape genomes has provided a foundation for accelerating prediction of gene families and identification of novel genes in other plant taxa. Different computational tools are available for analyzing and clustering the genetic information into various functional segments (gene families, TE, etc.) based in part on information from such reference genomes.

2.3.3 Synteny and Collinearity

The terms synteny and collinearity have been extensively used to denote the presence of two or more genes on corresponding chromosomes and in corresponding order, respectively. The extent of synteny and co-linearity vary among the flowering plant genomes largely due to recurring whole genome duplications followed by massive gene loss, with chromosomal rearrangements superimposed on post-duplication gene loss in breaking the ancestral gene ropes (Bowers et al. 2003b). Conventionally, synteny and co-linearity have been estimated by matching one-to-one (pair wise) conservation between the species. Recently, multiple way co-linearity analyses offer improved resolution for inferring ancestral gene order in angiosperms. For example, in *Oryza-Arabidopsis* comparison, the longest collinear segment contains 23 orthologous gene pairs. This is improved twofold to 47 genes by incorporating *Vitis* genome sequence into the comparison (Tang et al. 2008a).

After polyploidization in angiosperms, differential gene loss or retention is a very important factor affecting synteny between corresponding regions on different chromosomes (Paterson et al. 2006). All these properties coupled with variability in DNA substitution rates among plants, make rational of using shared co-linearity as a reliable phylogenetic character. Moreover, accurate synteny/collinearity among different genomes or subgenomes may help in elucidating the ancestral angiosperm genome which will then provide a universal reference point in understanding the nature and genes conferring phenotypic diversity across the angiosperms.

Instances of synteny and or collinearity have been detected in plant taxa that diverged from a common ancestor long ago. For example, cotton and *Arabidopsis,* diverged from a common ancestor about 83–86 mya (Benton 1993) revealed appreciable conserved gene order (Rong et al. 2005). Our ability to discern conserved gene order is not uniform across a genome – for example, pericentromeric regions that are gene-poor and in which repeat-associated rearrangements may persist a long time, are notoriously difficult in which to discern conserved synteny. Moreover, other genomic regions that went through structural changes may tolerate these changes with different frequencies, provoking tantalizing hypotheses for future study about whether these rapidly-evolving regions have a disproportionate impact on phenotypic divergence or even reproductive isolation between plant taxa.

2.3.4 Comparison of Sequenced Genomes

Comparisons among the sequenced genomes of *Arabidopsis*, rice, poplar, grapevine, papaya and sorghum shed light on ancient gene orders. For example, most papaya segments correspond to three or four *Arabidopsis* segments, which shows that two genome duplications have affected the *Arabidopsis* lineage since its divergence from *Carica* (Ming et al. 2008). Individual *Arabidopsis* genome segments correspond to only one *Carica* segment, revealing that *Carica* has not experienced duplication since its divergence from *Arabidopsis*. Both *Vitis* and *Carica* share only one duplication event, γ, with *Arabidopsis,* while α and β occurred within the *Arabidopsis* lineage (Tang et al. 2008a).

Large-scale (segmental or complete) genome duplication affects not only gene copy number but also substantially fractionates the ancestral gene linkages across multiple chromosomes. Recently, this problem was overcome using a robust computational framework that combines information from multiple orthologous and duplicated regions to construct local syntenic networks, which is instrumental in deducing the footprints of "paleo-hexaploid" before diverging from a common ancestor (Tang et al. 2008b). These findings will pave the way for approximating the number and arrangement of genes in the last universal common ancestor of angiosperms (Tang et al. 2008b).

2.3.5 Conserved Gene Position Does Not Necessarily Mean Conserved Function

Genome alignments permit us to deduce positional correspondence between genes or groups of genes, but provide no direct information about gene function. Indeed, whole-genome duplication and the associated adaptation of the new gene set to the duplicated state may be closely associated with changes in functionality, particularly of those genes that do persist in duplicate. Gene duplication often leads to over expression of some genes and silencing of others, as has been found in synthetic allopolyploids

of wheat and cotton (Adams et al. 2003; Kashkush et al. 2003). Retention of duplicated genes following polyploidy might be due to genomic buffering of dosage-sensitive genes (Chapman et al. 2006; Thomas et al. 2006), or to differential expression and/or sub-functionalization of other paralogs (Lynch and Force 2000).

Allopolyploidy addition to internal chromosomal rearrangements also changes the gene expression in different plant taxa including triticale (Chen 2007). Polyploidy is associated with a host of non-additive phenotypes that may be a result of changes in gene regulation and expression. For example, each of the allotetraploids *B. napus* (n = 19) (Udall et al. 2006) and *G. hirsutum* (Jiang et al. 1998) yield more than their diploid progenitors, possibly due in part to a relative increase in number of regulators, controlling gene expression in polyploids (Osborn et al. 2003). Genome-wide expression profiling in synthetic allotetraploid *A. thaliana* × *A. arenosa* has shown non-additive gene regulation (Wang et al. 2006). Delay in flowering time is attributed to epistatic interactions between two loci, one for flowering from *A. thaliana* and the other from *A. arenosa* (Wang et al. 2006). There is an important need to better understand the radical changes occurring after polyploidization, and their relationship to the evolution of genetic novelty.

2.3.6 Comparative Analysis of Small RNA

Small RNAs (siRNAs), important players in gene regulation, comprise 21–24 long nucleotide fragments, that have been most thoroughly studied in *Arabidopsis*. On the basis of their function and maturation, small RNAs have been categorized into short interfering RNAs (siRNAs) and micro-RNAs (miRNAs), both are involved in gene silencing by targeting complementary mRNA (Bartel 2004). The identification of siRNAs has largely been accomplished by comparative methods, which detect only conserved families of siRNAs (Rajagopalan et al. 2006). Deep sequencing of siRNAs permits the discovery of new and small classes of these molecules (Henderson et al. 2006).

There is remarkable variation in the size of miRNA populations across plant genomes, and the number that are species specific. For example, 114, 164 and 182 miRNAs has been discovered in *Arabidopsis*, grapevine and rice respectively (Griffiths-Jones 2004; Griffiths-Jones et al. 2006; Jaillon et al. 2007). Microevolution and regulation of miRNAs is an attractive subject for further investigation to explore the nature of miRNAs, their role in gene regulation, and rate of evolution in their target sites.

2.3.7 QTL Comparisons Across Taxa

QTL mapping experiments provide heterogeneous results due to the use of different genotypes, environments, and sampling variation. Compilation of QTL mapping results yields a more complete picture of the genetic control of a trait than can be

obtained in any single study, and reveals the genomic organization of trait variation. In an early investigation, QTL for domestication traits in independently-domesticated cereal lineages fell into corresponding intervals of the respective genomes more often than could be explained by chance (Paterson et al. 1995). Likewise, non-random correspondence of QTLs associated with domestication traits (e.g. fruit quality, weight and shape) is also found among tomato, pepper, and eggplant (Gephardt et al. 1991; Tanksley et al. 1992; Frary et al. 2000, 2003; Thorup et al. 2000; Doganlar et al. 2002a; Van der and Tanksley 2003).

The flowering time is conferred by numerous QTLs in maize (Chardon et al. 2004), and its molecular basis appears to be correlated across the maize–sorghum–rice lineages (Lin et al. 1995; Salvi et al. 2007) Similarly, conserved QTLs confer 'timing of bud burst' in oak and chestnut tree species (Casasoli et al. 2006) and this conservation may aid in the identification of these genomic regions. Non-conserved QTLs may have roles in local adaptation and species differentiation (Casasoli et al. 2006). A particularly extensive meta-analysis of 432 QTLs mapped in one diploid and ten tetraploid cotton populations revealed the sub genomic distributions of QTLs, identified hotspots for QTLs affecting particular traits, and showed that single-gene mutants in fiber characteristics may profoundly alter the QTL landscape, suggesting that fiber-related traits may be controlled by a complex gene network (Rong et al. 2007).

It is important to emphasize that non-random correspondence of QTL position does not necessarily mean that same genes confer the QTLs. For example, QTLs associated with the *tb1* gene conferring branching in maize, are conserved across the Poaceae family, but have a very meager role in foxtail millet (Doust et al. 2004). Such non-random QTL distributions could reflect non-random distributions of particular gene functional groups (i.e. gene clusters).

In isolating genes responsible for QTLs, microcollinearity aids in designing new markers from well-studied plant species to be applied to more complex genomes or orphan crops. For instance, two wheat QTLs associating with vernalization genes (*VRN1* and *VRN2*) (Dubcovsky et al. 1998; Tranquilli and Dubcovsky 2000) were cloned in part by designing additional markers based on the rice and sorghum genomes. Rice is syntenic to many corresponding regions containing genes involved in different developmental processes such as tassel and ear development in maize, and has been utilized for cloning these genes. Likewise, sorghum is closely related to major cellulosic biofuels crops including *Saccharum* (sugarcane) and *Miscanthus*, and its sequence is expected to accelerate progress in improvement of biofuels crops. Conservation of gene order between rice and barley genomes was not found in regions containing disease resistance (Bortiri et al. 2006). However, two of the three QTLs associated with clubroot resistance in *B. rapa* are syntenic to a disease resistance gene cluster, also termed as major recognition complexes (MRCs, Speulman et al. 1998) on chromosome 4 of *Arabidopsis* suggesting a possible common origin of MRCs from an ancestral genome (Suwabe et al. 2006).

Comparison of multiple QTL mapping studies is not only advantageous in molecular breeding experiments but also has increased utility in cloning of the genomic regions. A total of 432 QTLs conferring different traits in 11 cotton populations were

aligned and projected on a reference map which fosters deducing synteny between the cotton and *Arabidopsis* genomes, revealing non-random distribution of fiber and trichome-related genes in the two genomes (Rong et al. 2007). This correspondence suggests that the identification of genes for trichome development in *Arabidopsis* may lead indirectly to the identification of fiber-related genes in cotton.

2.3.8 Defining Unification in Biological Terminology

Vavilov's (1922) law of homologous series in variation foreshadowed the DNA-level demonstration of synteny across the genomes of different plant species. More recently, microsynteny studies revealed co-linearity at the level of genes between different species belonging to the same family. An early comparative genetic map developed among members of Solanaceae family highlighted the conserved order of DNA markers among genomes with two- to fourfold size variation (Grube et al. 2000). Similarly, positional conservation of gene order among the members of family Poaceae using cDNA anchor probes has been reported by numerous investigators.

Briefly, comparative genetic mapping has shown extensive co-linearity in different genomes belonging to the same family. Thus, the exploitation of genome co-linearity can help in determining correspondence among chromosomes of species belonging to the same tribe or family, which will be instrumental not only in categorizing chromosomes on the basis of sharing common genetic information but also in inferring the set of chromosomes of an ancestral angiosperm genotype.

2.4 Comparative Genomic Studies on Major Crops

Pre-genomic discovery of parallel arrangement in morphological and isozyme markers among different taxa rationalized the continuation of such studies at the DNA level, revealing significant co-linearity across even divergent plant taxa (Bowers et al. 2005; Salse et al. 2008; Tang et al. 2008a). Large-scale genome sequencing efforts coupled with strong computational algorithms have provided a detailed view of the complex evolutionary forces that shaped the structure of eukaryotic chromosomes (Eichler and Sankoff 2003) which are discussed in the following sections.

2.4.1 Poaceae

In multiple investigations, conservation in gene order and content were reported among the Poaceae (Goff et al. 2002; Sorrells et al. 2003; Bowers et al. 2005; The Rice Chromosome 3 Sequencing Consortium 2005; Zhu and Buell 2007; Salse et al. 2008).

Genetic maps based on RFLP markers revealed significant macrocollinearity between the cereal genomes (Ahn and Tanksley 1993; Paterson et al. 1995; Devos and Gale 2000; Feuillet and Keller 2002; Devos 2005), and established inter-relationship among the homeologous chromosomes of the Triticeae and their relatives (Naranjo et al. 1987; Chao et al. 1989; Namuth et al. 1994; Devos et al. 1995; Hohmann et al. 1995; Marino et al. 1996; Mickelson-Young et al. 1995; Nelson et al. 1995; Van Deynze et al. 1995; Ming et al. 1998; Paterson et al. 2000, 2004; Salse et al. 2008). Comparative mapping using RFLP markers often probed paralogous rather than orthologous sequences, leading to underestimation of collinearity using these markers.

The availability of genomic resources such as drafts of the rice and sorghum genome sequence and huge collections of ESTs of rice (http://www.gramene.org), sorghum (http://cggc.agtec.uga.edu), wheat (http://www.ncbi.nlm.nih.gov/dbEST), maize (http://www.panzea.org) and many more, has made it possible to compare genome sequences with non-sequenced plant taxa (Ming et al. 2008; Tang et al. 2008a) to refine our understanding of collinearity, genetic diversity and agricultural productivity (Paterson et al. 2003, 2005). For example, the sequences of rice chromosomes 3 and 11 were compared with wheat ESTs (Singh et al. 2004) to increase the resolution of comparative mapping. Maize derived ESTs were compared with the sequence of rice chromosome 3 (Buell et al. 2005), which showed relatively more rearrangements in the cereal genomes.

Whole genome duplication (WGD) occurred before the divergence of cereal genomes (Paterson et al. 2004), supported by RFLP markers in rice (Kishimoto et al. 1994; Nagamura et al. 1995) and sorghum (Chittenden et al. 1994), and the complete genome sequences of the japonica and indica subspecies of rice (Goff et al. 2002; Yu et al. 2002, 2005; Paterson et al. 2003, 2004; Vandepoele et al. 2003; Guyot et al. 2004; International Rice Genome Sequencing Project 2005; Wang et al. 2005).

A large scale comparative genomic analysis of rice with four plant genome sequence data sets including *Arabidopsis* (The *Arabidopsis* Genome Initiative 2000), poplar (Tuskan et al. 2006), sorghum (Bedell et al. 2005) and maize (Palmer et al. 2003; Whitelaw et al. 2003), and transcript assemblies from 185 plant species, confirmed 38,109 (89.3%) of the total 42,653 non-transposable element-related genes. Out of these, 7,669 genes are lineage-specific which may have a role in species diversity (Zhu and Buell 2007). The conservation of short intergenic regions in a few genes is attributable to alternative splicing (Wang and Brendel 2006), a common phenomenon in the genome of rice, sorghum and maize. However, its role in biological function is unknown.

In the maize genome, duplications were reported by comparing the orthologous loci between rice, sorghum and maize (Swigonova et al. 2004). In another study, comparison of 2,600 maize mapped sequence markers with the rice genome sequence revealed six duplications between the maize and rice chromosomes (Salse et al. 2008) which has been supplemented by developing a high resolution comparative physical map between rice and maize genomes (Wei et al. 2007). However, the mechanism of duplications, and discovery of the exact number of basic chromosomes of the grass ancestor, is big challenges for researchers. There is inconsistency in the literature about the exact number of ancestral grass chromosomes. Two

models have been proposed, exhibiting 5–12 (Gaut 2002; Wei et al. 2007) and five (Salse et al. 2008) chromosomes in the ancestral grass genome.

The sorghum genome has not incurred genome duplication since the pan-cereal event (Paterson et al. 2004), making it an excellent out group for comparison with the duplicated genomes of maize and sugarcane, the worlds leading biofuel crop (Bedell et al. 2005; Bowers et al. 2005). Comparative mapping of *Sorghum propinquum* (a closely related species) with sorghum and *Oryza longistaminata* shed light on genes related to domestication (Paterson et al. 1995; Hu et al. 2003). A comparative physical map of sorghum with rice and maize has identified probable euchromatic regions in the grass family (Bowers et al. 2005).

A consensus map of *Saccharum* species and sorghum reveals a high degree of conservativeness between the two genomes, which can be instrumental in translating the genetic information for developing a detailed genetic map (Ming et al. 2002). Complexity and variable chromosome content of the sugarcane genome are major concerns in its whole genome sequencing (Dillon et al. 2007). Comparison of sugarcane genetic maps and EST sequences are the major tools utilized for understanding its genome, and dissecting traits of interest. Out of 237,954 ESTs derived from *S. officinarum*, two-thirds have recognizable homologs in *Arabidopsis*, and the remainder appear to be monocot specific or perhaps rapidly-evolving (Vincentz et al. 2004). Additional EST sequences are needed for comparison with the model species to identify genes of potential agronomic value (Vettore et al. 2003) and new EST-SSRs (Cordeiro et al. 2001) will be useful for developing detailed genetic linkage maps.

The Triticeae tribe diverged from rice around 50 mya (Paterson et al. 2004) and contains *Triticum aestivum* species, one of the most important food crops, which originated around 10,000 years ago through hybridization of diploid ancestors (A, B and D genomes) (Feldman et al. 1995). Over its short evolutionary span, genes remain conserved in comparison with those of *T. monococcum* (derived from *T. urartu*, the A genome progenitor of present day wheat), however, the intergenic regions have gone through several changes (Wicker et al. 2003; Isidore et al. 2005; Gu et al. 2006). Extensive conservation in intergenic regions was also discovered between *T. aestivum* and *T. turgidum* (spp. durum), which diverged approximately 800,000 years ago (Gu et al. 2006). In another investigation, Pm_3 loci isolated from hexaploid wheat (Yahiaoui et al. 2004, 2006), conferring resistance to powdery mildew, have gone through an extremely dynamic evolution which resulted in minimal sequence conservation between the A genome species at three ploidy levels. This wheat locus evolved more rapidly than its homolog in rice (Wicker et al. 2007).

2.4.2 Solanaceae (Tomato; Potato; Eggplant; Peppers)

The members of the Solanaceae family especially tobacco (*Nicotiana tabacum*), potato (*S. tuberosum*) and eggplant (*S. melongena*) have relatively few rearrangements in their genomes. Low-resolution comparative mapping suggests that disease

resistance loci (R) are well conserved in the Solanaceae (Grube et al. 2000; Pan et al. 2000) while evolving much faster in the crucifer and grass genomes (Gale and Devos 1998; Leister et al. 1998; Paterson et al. 2000). The genomic region in the long arm of tomato chromosome 11 has been conserved across many members of the Solanaceae, and has been investigated for isolation of new R genes using comparative genomic approaches (El-Kharbotly et al. 1994, 1996; Lefebvre et al. 1995; Ori et al. 1997; Simons et al. 1998; Hanson et al. 2000). The R_3a gene (confers race-specific resistance to the late blight pathogen) was isolated from potato which corresponds with the I_2 complex locus in tomato. After the divergence of potato and tomato, the R_3 locus has gone through a significant expansion in number of genes and functional diversification while keeping intact the flanking co-linearity (Huang et al. 2004, 2005).

In multiple investigations, different degrees of restructuring within and across plant taxa was reported (Lagercrantz and Lydiate 1996; Livingstone et al. 1999; Ku et al. 2000; Thorup et al. 2000; Fulton et al. 2002; Salse et al. 2002; Bennetzen and Ma 2003; Nishiyama et al. 2003). For example, within Solanaceae, the genomes of potato and tomato differ only by five paracentric inversions (Bonierbale et al. 1988; Tanksley et al. 1992; Doganlar et al. 2002a), suggesting a high propensity for intrachromosomal rearrangements. Around 30 chromosomal breaks distinguish the genomes of tomato and pepper (Prince et al. 1993; Livingstone et al. 1999), a relatively high degree of chromosomal rearrangements. Genomic comparison with other plant taxa have shown a syntenic relationship of 90 blocks covering 41% of the potato genetic map with 50% of the *Arabidopsis* physical map, indicated that ancient intra- and interchromosomal duplications has played a role in potato genome evolution (Gebhardt et al. 2003).

2.4.3 Fabaceae (Leguminosae)

This is the third largest family of angiosperm (650 genera and 18,000 species, Polhill and Raven 1981), contributing 27% of world food production in the form of edible oil, food and feed (Graham and Vance 2003). Biological nitrogen fixation is one of the distinctive features of the legumes (Dixon and Sumner 2003). In early genetic mapping studies, conserved syntenic blocks were observed among the various legumes (Menancio-Hautea et al. 1993; Boutin et al. 1995; Lee et al. 2001). Useful degrees of gene linkage conservation have been reported between lentil and pea genomes (Weeden et al. 1992); and between mung bean and cowpea (Menacio-Hautea et al. 1993). Use of anchor markers, designed on the basis of unique gene sequences (Choi et al. 2006), in multiple legume species revealed extensive synteny between model and crop legume species (Zhu et al. 2005; Choi et al. 2006).

Early pioneering studies suggested a simple genetic relationship between *P. sativum* and *Lens culinaris* (Weeden et al. 1992). Subsequent comparative genetic studies using sophisticated DNA markers have clearly shown macrosynteny between lentil (n = 7) and *M truncatula* (n = 8), and also detected syntenic relationships

among important domestication traits such as pod indehiscence, flower color, seed coat pattern, and *Fusarium* wilt between the two species. Moreover, moderate level of chromosomal rearrangements might be the reason for differences in chromosome numbers between the two species (Phan et al. 2007).

Limited genomic studies have been conducted on the genus *Arachis* because of its monophyletic polyploid origin, resulting in little genetic diversity (Burow et al. 2001). Polyploid speciation triggered significantly the genome restructuring (Song et al. 1995; Chen et al. 1998), and may also have influenced *Arachis* species (Burow et al. 2001). Genetic variations in repetitive elements in the genomes of diploid progenitors may help to drive speciation in *Arachis* (Seijo et al. 2007). Genome duplication predating the divergence of the diploid progenitors (A&B) has been suggested (Burow et al. 2001).

In legumes, two model species *Medicago truncatula* and *Lotus japonicus* are being largely sequenced (Young et al. 2005), which will be of particular interest in cloning of genes involved in symbiosis (Schauser et al. 1999; Endre et al. 2002; Krusell et al. 2002; Nishimura et al. 2002; Stracke et al. 2002; Madsen et al. 2003; Ané et al. 2004; Levy et al. 2004). Partial genome sequences of 100 Mb each for *L. japonicus* and *M. truncatula* revealed a strong conservation of genome structures (Cannon et al. 2006). Similarly, syntenic blocks of as large as whole chromosome arms have been detected between the chromosomes of these two species based on genetic mapping (Young et al. 2005). Low levels of intragenomic synteny within each of the genome indicated the occurrence of ancient whole genome duplication followed by gene loss and chromosomal changes (Cannon et al. 2006). Conservation between *M truncatula* and *L japonicus* genomes is higher than that of *M truncatula* and *Glycine max* (Ku et al. 2000; Choi et al. 2004).

In other studies, a high level of macrosynteny between the *Pisum sativum* and *M. trunctula* genomes and long tracts of macrosynteny between *M. truncatula* and *Phaseoleae* species (soybean and mungbean) enabled the construction of a genome-wide picture of legume synteny in the form of concentric circles of corresponding chromosomes anchored by *M. truncatula* (Choi et al. 2004).

A number of microsynteny studies between the closely related model species (*M. trunctula* and *L. japonicus*) and crop species (pea) enabled positional cloning of symbiosis related genes (Endre et al. 2002; Stracke et al. 2002), and have been stretched to the more evolutionarily distant soybean showing appreciable collinearity and microsynteny among BAC contigs (Yan et al. 2003, Mudge et al. 2005). Remnants of synteny were also identified between *G. max* and *A. thaliana* (Grant et al. 2000; Lee et al. 2001). An extended sequences comparison of the tandemly duplicated N-hydroxycinnamoyl/benzoyltransferase (HCBT) gene family derived from *G. max*, *M. truncatula* and *A. thaliana* revealed a network of synteny within conserved regions, interrupted by gene loss and rearrangements (Schlueter et al. 2008). Comparison of model genomes with the forthcoming soybean genome (Jackson et al. 2006) will help in further clarifying gene order and synteny among the different genomes of the family Fabaceae.

2.4.4 Brassicaceae

The family Brassicaceae, also called the "mustard family" contains the most important model species *A thaliana* and also many important agricultural species including three diploids (*B. rapa*, *B. nigra*, and *B. oleracea*) and three tetraploids (*B. juncea*, *B. napus* and *B. carinata*) (Schranz et al. 2006). Comparative mapping in Brassicaceae has been active since the development of genetic linkage maps using DNA markers in the early 1990s (Karp 1998) for major crop species of the genus *Brassica* (Slocum et al. 1990; Landry et al. 1991; Song et al. 1991; Chyi et al. 1992; Kianian and Quiros 1992; Figdore et al. 1993; Ferreira et al. 1994; Kowalski et al. 1994; Teutenico and Osborn 1994; Uzunova et al. 1995; Lagercrantz and Lydiate 1996; Truco et al. 1996; Lan et al. 2000). These maps showed substantial collinearity interrupted by small inversions (Kianian and Quiros 1992; Lan et al. 2000). Extensive genome changes in synthetic polyploids fueled the debate on rapid genome changes following polyploidization (Song et al. 1995) which set the stage to explore this mechanism (Osborn 2004; Soltis et al. 2004).

The role of *A. thaliana* (n = 5) in building a Brassicaceae genomics circle is marginal because of its derived nature. A total of 38% of the species contain chromosome number n = 8 (Warwick and Al-Shehbaz 2006) and several chromosomal rearrangements are unique to *A. thaliana* (Lysak et al. 2007). The two species *C. rubella* and *A. lyrata* (Kuittinen et al. 2004; Yogeeswaran et al. 2005) containing n = 8 may be better reference points in comparative genomic studies based on sharing genomic collinearity, the conservativeness of genomic blocks (Boivin et al. 2004; Kuittinen et al. 2004; Yogeeswaran et al. 2005) and similar genome structure (Koch and Kiefer 2005) with *A. thaliana*. Through comparative chromosome painting, the pattern of chromosome specific BAC contig probes derived from *A. thaliana* suggested a common mechanism (pericentric inversion followed by reciprocal translocation/fusion event) of chromosome number reduction in all the species of Brassicaceae (Boivin et al. 2004; Kuittinen et al. 2004; Yogeeswaran et al. 2005) and position of centromeres in the ancestral karyotype (Lysak et al. 2006), confirmed by genetic mapping in *A. lyrata* (Hansson et al. 2006; Kawabe et al. 2006).

Despite the aforementioned difficulties in comparing the genome of *Arabidopsis* with *Brassica* species (Koch et al. 2007), 21 conserved syntenic blocks making 90% of the genome of *B. napus* (n = 19) were found collinear, having been maintained since the divergence of the *Arabidopsis* and *Brassica* lineages around 20 mya (Kowalski et al. 1994; Yang et al. 1999; Koch et al. 2003; Lysak et al. 2005). In another study, sequences of a set of five genes (located on a 15 kbp segment of *A. thaliana* chromosome 3) in *B. rapa*, *B. oleracea* and *B. nigra* were physically closely linked (Sadowski et al. 1996). A 30 kbp segment of *A. thaliana* chromosome 4 harboring six genes were found in comparable organization in the corresponding regions in *B. nigra*, however, this region of the genome was considerably larger in *B. nigra* than in *A. thaliana* (Sadowski and Quiros 1998). The *S* locus region from *B. campestris* and *A. thaliana* revealed extensive collinearity in the homeologous region (Conner et al. 1998).

Later, an ancestral karyotype containing 24 conserved blocks (n = 8) was drawn based on comparative maps between *Arabidopsis* and two Camelineae species n = 8 (*Arabidopsis lyrata* and *Capsella rubella*, Lysak et al. 2006) which was aligned with the 21 conserved blocks in *Brassica* to infer the ancestral karyotype (Schranz et al. 2006). These regions share a common mechanism of genome evolution. Moreover, macrosynteny between *Boechera stricta* (n = 7) and Brassicaceae species helps in clarifying the chromosome number reduction in the genus *Boechera* (Schranz et al. 2007).

2.4.5 Malvaceae

The Malvaceae comprises three main species-rich lineages including Gossypieae (n = 10 to 13, Fryxell 1968), Malveae (n = 5 to 21, Bates 1968; Bates and Blanchard 1970) and a group of taxa belonging to the large genus *Hibiscus* (n = 12 to around 144, while n = 14 or 18 is the most prevalent, Fryxell 1968, 1988; Pfeil et al. 2002, 2004). In comparative genomic investigations, the genus *Gossypium* belonging to the tribe Gossypieae, received much attention because of its paleopolyploid nature (Rong et al. 2005), and is the leading natural fiber producer. The genus *Gossypium* includes five allopolyploid and 45 diploid species (Fryxell 1979, 1992) containing eight genomes designated A through G and K (Beasley 1940; Phillips and Strickland 1966; Edwards and Mirza 1979; Endrizzi et al. 1985).

The allotetraploid cotton species originated through uniting the A and D genomes about 1–2 mya (Wendel 1989; Wendel and Albert 1992). Comparative maps developed through surveying RFLP markers on four cotton genomes (A, D, Dt and At) showed conservation of gene order and synteny (Reinisch et al. 1994; Brubaker et al. 1999). Later, detailed STS-based genetic maps reiterated the collinearity between the A and D genomes (Rong et al. 2004). The A diploid genome can be differentiated from the At genome (representing the diploid A genomes in the tetraploid cotton species) by two reciprocal translocations occurred only in At genome following allotetraploid formation; and one inversion (Brubaker et al. 1999; Rong et al. 2004), complementing earlier findings based on conventional cytogenetic approaches (Brown 1980; Menzel et al. 1982). Similarly, five inversions identified while comparing the maps of the four genomes, can differentiate the D and Dt genomes.

Ancient genome duplication is a very common phenomenon in the evolution of many plants. The *Arabidopsis* and cotton lineages each have incurred ancient duplication after their divergence, although retaining discernible correspondence in genetic loci (Rong et al. 2005). The second tribe of the family Malvaceae is "Malveae," comprised of the members of the genus *Hibiscus* (Pfeil et al. 2002) exhibiting a wide range in basic chromosome number (Fryxell 1968, 1988) with a basic haploid set ranging from n = 5 to 21 (Bates 1968; Bates and Blanchard 1970). Like other crop plants, the status of *Hibiscus* polyploids was studied by exploring a single copy gene rpb_2 (Oxelman et al. 2004) which was duplicated before the divergence of the *Hibiscus* genus (at the subfamily level). Moreover, after duplication,

both copies of the gene retain their function and structure in this lineage, a compelling factor motivating selection of this gene in probing relatively old duplication events and in phylogenetic analysis (Pfeil et al. 2004).

2.5 Flow of Genetic Information from Well-Studied to Less Explored Genomes

The evolution of plants from a common ancestor left collinearities and syntenies in genome structure and function, the fundamental framework to transfer genetic information from model species to less studied relatives. In flowering plants, an important role of model plant species is to serve as a source for transferring genetic knowledge to crop plants via genome sequences and genetic linkage maps (Sato et al. 2007). Among the plant taxa, Brassicaceae, Solanaceae, and Poaceae include botanical models which can be distinguished on the basis of nature and number of rearrangements from their less studied relatives (Bonierbale et al. 1988; Ahn and Tanksley 1993; Prince et al. 1993; Periera et al. 1994; Gale and Devos 1998; Doganlar et al. 2002a), which may lead to reproductive isolation and ultimately speciation (Rieseberg 2001).

The Brassicaceae (or Cruciferae) is a dicot family comprises of 360 genera (Al-Shehbaz 1973) including the agronomically important genus *Brassica* and the model species *A. thaliana*. The complete genome sequence information of *A. thaliana*, detailed information of its physical map position, copy number of genes and intergenic sequences, location of duplicated chromosomal segments (The *Arabidopsis* Genome Initiative 2000; Blanc et al. 2000; Vision et al. 2000; Bowers et al. 2003b; Ermolaeva et al. 2003; Raes et al. 2003), and its small genome size provides a foundation for comparative mapping studies (Yogeeswaran et al. 2005). *Arabidopsis thaliana*, the crown botanical model, has now been suggested to actually be a relatively poor model because of its genomic complexity. For example, *Arabidopsis* which is taxonomically closest to papaya among the sequenced genome exhibited two genome duplications since its divergence from *Carica* (Ming et al. 2008).

Rice is one of the most important plant species of the family Poaceae, and is a genetic model not only for the other members of this family but also for the monocotyledonous plants. The map-based sequence of the rice genome (*O. sativa* ssp. *japonica* var. Nipponbare) and a draft sequence of the *indica* subspecies (var. 93–11) (Yu et al. 2002, 2005) have been used in annotating genes and gene prediction and deducing evolutionary consequences in other cereals and grass species (Yuan et al. 2005; Tang et al. 2008a).

In the Solanaceae family, tomato (*Solanum lycopersicum*) has been investigated both by classical and molecular genetic procedures (Tanksley et al. 1992; Livingstone et al. 1999), facilitated by its diploid (2n = 2x) genome of 12 chromosomes containing long stretches of euchromatin at the distal ends and heterochromatic regions flanking the centromeres (De Jong 1998; Kulikova et al. 2001; Fransz et al. 2003). Tomato served as a model system for studying fruit development and ripening because of its rich genetic resources including both germplasm and mutants; a dense molecular map;

comprehensive EST dataset (http://www.sgn.cornell.edu/); and cDNA microarray (Giovannoni 2004); facilitating comparative genomic studies within and across taxa.

Among the 50 species in the genus *Gossypium*, the *G. raimondii* 'D' genome has been prioritized for genome sequencing because of its relatively small genome size and minimal repetitive DNA. The generated information will help in filling the major void in international *Gossypium* genomics resources, paving the way for dissecting the genetic control of specific *Gossypium* traits. The first *Gossypium* genome will also foster tests of hypotheses about features of *Gossypium* genome organization and/or evolutionary history that may be related to adaptive features that permit different *Gossypium* taxa to flourish in a wide range of environments, and also to be adapted to agriculture (Chen et al. 2007).

2.6 Evolutionary Consequences

Evolution of plant species is the outcome of numerous processes including gene duplication and loss, whole genome duplication, fluctuation in ploidy level, retrotransposon activity, and genome rearrangements (Goff et al. 2002; Yu et al. 2002; Wicker et al. 2003; Kellogg and Bennetzen 2004; Caicedo and Purugganan 2005; Ming et al. 2008). Trends in genome size fluctuate substantially in different plant lineages. Expansion of the sorghum genome relative to rice occurred largely in heterochromatic regions comprised largely of LTR-retrotransposons (Paterson et al. 2009). A total of 58% of the sorghum genome is comprised of pericentromeric heterochromatin versus 15% of rice. The turnover of cereal genomes retrotransposons has been rapid in cereal crops (Swigonova et al. 2005).

Through comparative sequence analysis, wide range in genome size variation in *Gossypium* genome has largely been attributed to differential proliferation of dispersed repetitive elements (Zhao et al. 1998; Hawkins et al. 2006) which was detected in heterochromatic regions located between highly conserved, euchromatic gene islands (Grover et al. 2004; Hawkins et al. 2006). Genome size fluctuation may also be affected by other mechanisms including variation in intron length, expansion/contraction of tandem repeats, illegitimate recombination, indel bias, and unequal intrastrand homologous recombination (Petrov and Wendel 2006). It was also speculated that smaller genomes carry smaller introns (Deutsch and Long 1999; Vinogradov 1999); however, such a correlation was not observed in the genome *Gossypium* (Wendel et al. 2002; Grover et al. 2004).

2.6.1 Phylogenomics: Contributions to Developing a Consensus Tree

Variations in gene sequences among different plant taxa have been instrumental in constructing the phylogenetic tree of the plant kingdom and also estimating the molecular clock of plant phylogeny (Miller et al. 2004; Caicedo and

Purugganan 2005). Different nuclear introns, exons, ribosomal genes, and plastid gene sequences have increasingly been surveyed for estimating genetic diversity and phylogenies of plant species for which genome sequence information is not available (Cronn et al. 2002; Shaheen et al. 2006; Ahmad et al. 2007). DNA markers including SSRs, RAPD, AFLP and SNPs have also been used for estimating genetic diversity, and understanding the phylogenetic relationship among major plant taxa (Abdalla et al. 2001; Mukhtar et al. 2002; Rahman et al. 2002, 2008).

In distantly related species, if whole genome sequence information is not available, nuclear genes, preferably single copy genes (Qiu et al. 1999; Fulton et al. 2002), and mitochondrial and plastid gene sequences (Qiu et al. 1999; Palmer et al. 2004) were used to construct the tree of life. For example, in Brassicaceae, 338 genera comprising of 3,700 species were classified into 25 tribes (Beilstein et al. 2006) using nuclear (Bailey et al. 2006) and chloroplast based markers (Beilstein et al. 2006).

Sequences of the *Arabidopsis*, rice, sorghum, papaya, grape and poplar genomes provide reference points to initiate and/or deepen insight into the complex genomes for drawing a comprehensive phylogenetic relationship of flowering plants. In the future, proliferation of genomic data will help in constructing the comprehensive and well adapted phylogenetic tree and also further exploring molecular clocks (Miller et al. 2004).

2.6.2 Direct Implications in Gene Loss/Gain for Attaining New Functions

The recurring whole genome duplications in angiosperms over the last 200 million years followed by gene loss and mobile element activities yielded many novel plant taxa (Bowers et al. 2003b; Tang et al. 2008a). However, the number and size of gene families are similar in the well studied genomes such as sorghum, *Arabidopsis*, rice and poplar (Paterson et al. 2009). Around 58% of gene families were shared among all four species; 24% were monocot specific and 7% were unique to sorghum, a fraction which is similar to *Arabidopsis* (6.7%, The *Arabidopsis* Genome Initiative 2000). However, fewer gene families are specific to rice (3.7%, Matsumoto et al. 2005) while many are unique to poplar (15.7%, Tuskan et al. 2006). A high number of genes being orthologous between the rice and sorghum suggests that most gene loss occurred in a common rice ancestor (Paterson et al. 2009). Lineage specific genes may have been derived from non-functional DNA sequences (Itoh et al. 2007); genes with high divergence rates and independent gene deletions among plant taxa are the major evolutionary forces for making genes unique to a particular species (Salzberg et al. 2001; Stanhope et al. 2001). In addition to the genomic changes, the impact of natural selection on favoring certain genes significantly contributed in shaping the genomes of plant species including papaya (Itoh et al. 2007; Ming et al. 2008).

Every plant species evolved new functions to sustain their survival, for example, papaya and poplar have more genes than *Arabidopsis* which are involved in cell expansion, larger plant size and lignin biosynthesis for evolving treelike habits. An abundance of genes associated with volatile features in papaya is the outcome of strong natural selection for enhanced attractants which may facilitate fruit dispersal by animals (Ming et al. 2008). Similarly, cytochrome P450 domain-containing genes that scavenge toxic compounds accumulated in response to stresses, are more abundant in sorghum (326, Paterson et al. 2009) than rice (228). There are many other gene families including expansin domain, kafirin (Song et al. 2004), NBS-LRR, and carbonic anhydrase (*cah*) gene families that fluctuate in their numbers among different crop species (Paterson et al. 2009).

2.6.3 Domestications of New Forms

A variety of factors including genetic architecture exerts impact on the ease and tempo of evolution. Domestication in many of the angiosperms is driven by changes at small numbers of loci with larger impacts (Cai and Morishima 2002). In contrary to this, domestication of sunflower has been preceded by selection on a large number of loci which confer moderate to small phenotypic impacts (Burke et al. 2002; Wills and Burke 2007). Domestication related genes are often found in clusters (Weeden 2007). It is suggested that species containing favorable genes in clusters were easily and rapidly domesticated, or alternatively the apparent gene clustering may be due to pleiotropic effects of individual genes (Doganlar et al. 2002a).

The domestication-associated traits act both in recessive and non-recessive manner. The recessive alleles presumably evolved because of "loss-of-function", for example, the nonshattering trait in cereals (Li et al. 2006), fruit weight trait in tomato and eggplant (Frary et al. 2000; Doganlar et al. 2002a), flowering time (Foucher et al. 2003) and dwarf habit (Martin et al. 1997) in pea, transition from two-rowed to six-rowed barley (Komatsuda et al. 2007), and other examples. Comparative analysis of domestication traits in three major cereal crops (sorghum, rice and maize) revealed correspondence in these genes/QTLs (Paterson et al. 1995). These genes can be cloned by adopting map-based cloning (Paterson 2002); association mapping (Buckler and Thornsberry 2002); and/or large-scale genomic scan in a crop species and its wild progenitor (Ross-Ibarra et al. 2007) depending on the crop species and available genetic information.

2.6.4 Polyploidy: Evolution of New Genomes

Polyploidy has contributed significantly to enhancing the productivity of plant taxa such as cotton, wheat, sugarcane, oat, soybean, peanut, canola, tobacco, coffee, and banana. The increase in genomic content of nuclei either brought about through

whole genome duplication or polyploidy (uniting of two genomes in one nucleus), followed by extensive chromosome restructuring (Brubaker et al. 1999) possibly through retrotransposon activity (Zhao et al. 1998), is one of the major causes of deviating from the arithmetic sums of the sizes of the parental genomes (Chen 2007). Presently, all the diploids including rice and *Arabidopsis* are apparently ancient polyploids (Bowers et al. 2003b). In ancient polyploids, domesticated forms evolve through tinkering with genes in their wild relatives (Doebley 2006). However, in recent allopolyploid species such as wheat, a null mutation of the GPC-B1 gene in the B genome, greatly facilitated by repetitive elements in the wheat genome (Flavell et al. 1974), caused a few days difference in the maturity (Slade et al. 2005). In the success of polyploids, the role of repetitive elements in impacting gene content still needs research and development efforts.

The reactivation of mobile elements (Kashkush et al. 2003; Madlung et al. 2005) resulting in chromosomal abnormalities, deletions, pollen sterility, and other problems may limit the chances of evolving successful synthetic allopolyploids such as allotetraploid *Arabidopsis* (Madlung et al. 2005) and wheat (Shaked et al. 2001). The process of accumulating chromosomal changes, especially deletions, is not sudden but is slow in examples such as the D genome of *T. aestivum* (Dvorak et al. 2004).

Cultivated cotton species *G. hirsutum* and *G. barbadense* evolved from a common paleopolyploid ancestor that formed about 1–2 mya by hybridization of an A genome species resembling *G. herbaceum* and a D genome species resembling *G. raimondii*. The two progenitors of tetraploid cottons are themselves thought to have shared a common ancestor about 6–11 mya (Wendel and Cronn 2003). The genus *Gossypium* is especially well suited to investigating the evolutionary consequences of polyploidy, as both the progenitor genomes of the allotetraploids are known and remain extant. The Dt genome (version of the D genome present in the allotetraploid) remains largely unchanged from that of its diploid progenitor, while two reciprocal translocations along with several inversions distinguish A from At chromosomes. While the diploid D-genome species does not produce spinnable fiber, it imparts important genes or regulators involved in fiber morphogenesis and its properties (Jiang et al. 1998; Paterson et al. 2003; Ulloa et al. 2005; Rong et al. 2007). Conversely, A-genome progenitor diploid species are extremely resistant to cotton leaf curl disease, while its derived tetraploid cotton species are susceptible (Rahman et al. 2005). Indeed, the A-genome species is more tolerant to many biotic and abiotic stress than AD-tetraploids. The loss of either A or D subgenome genes in subsequent generations by the AD-tetraploids, a phenomenon that is nearly ubiquitous among paleopolyploids but has only been studied to a limited degree in cotton, needs further investigation.

Most if not all Poaceae shared a common genome duplication (Paterson et al. 2004), with additional subsequent duplications in some lineages such as maize, wheat, and sugarcane. Despite this variation in ploidy and also substantial variation in genome size, chromosome number has remained more stable (Bennett and Smith 1991). Indeed, some polyploidizations have most probably been followed by a substantial degree of chromosome condensation, uniting two chromosomes into

one. For example *Z. mays* has been through one whole genome duplication (Swigonova et al. 2004) since the divergence of the *Zea* and *Sorghum* lineages – however, *Z. mays* contains the same chromosome number (n = 10) as *S. bicolor*, although the latter has not incurred genome duplication since the event shared by all cereals (Paterson et al. 2004). Likewise, some members of the Sorghum genus appear to have condensed from n = 10 to n = 5. Condensation of maize chromosomes would be consistent with the observation that single arms of several maize chromosomes correspond with entire sorghum chromosomes (Whitkus et al. 1992; Bowers et al. 2003a).

Building on a growing understanding of ancestral genome duplications and polyploidization events in the present day diploid cereals, low copy RFLP markers permitted the construction of a consensus grass map of low resolution based on 25 rice linkage blocks (Feuillet and Keller 2002). More recently, comparison of the rice genome sequence with wheat and maize cytological maps, have shown more chromosomal rearrangements between the genomes of rice, maize and wheat species, and also revealed evidence of ancestral genome duplications occurred between 53 and 94 mya (Paterson et al. 2003, 2004; Buell et al. 2005; Yu et al. 2005; Singh et al. 2007). All this information led to models proposing the basic chromosome number of the grass ancestor, ranging from n = 5 to 12 (Wei et al. 2007), and more recently n = 5 has gained support (Salse et al. 2008).

The basic chromosome set of wheat, barley and oat is x = 7, and one chromosome corresponds to the rice linkage groups 5 and 10 and another corresponds to a novel combination of parts of rice chromosomes 4 and 7 (Kellogg 1998). The members of tribe Bambuseae are polyploid except *Chusquea talamancensis* and possibly *C. subtesselata* (Judziewicz et al. 1999), and vary substantially in morphology perhaps in part due to passing through major genomic changes both in structure and function following polyploidy. It is hypothesized that the woody bamboos have two full copies of a rice-like genome which needs further investigation. *Zizania aquatica*, a wild relative of rice, has 15 chromosomes, with 14 that are clearly collinear with 11 of the 12 rice chromosomes. Moreover, three chromosomes seems to be duplicates of rice chromosomes 1, 4 and 9 (Kennard et al. 1999).

Sugarcane, an important biofuel crop, has undergone at least two genome duplications since its divergence from sorghum 8–9 mya (Jannoo et al. 2007). Polyploidy further added to the complexity of its genome (n = 18–85 versus n = 10 for sorghum) which not only impedes breeding progress but also makes challenging the job of understanding the sugarcane genome or its transmission genetics (Ming et al. 2005). Comparative genome mapping studies suggest a high degree of collinearity between the gene order of sugarcane and sorghum genomes (Ming et al. 2002).

In the Brassicaceae, three rounds of ancestral genome duplications (Vision et al. 2000; Bowers et al. 2003b) coupled with the polyploid origin of 37% of species playing a decisive role in the evolution of different plant taxa in this family (Warwick et al. 2006). Several comparative mapping studies between the *A. thaliana* and *Brassica* species have shown the footprints of triplication in numerous homeologous regions within *Brassica* (Lagercrantz 1998), complemented by sequencing (Town et al. 2006; Yang et al. 2006) and cytogenetic approaches

(Lysak et al. 2006; Ziolkowski et al. 2006), cumulatively supporting the possibility of having a common hexaploid ancestor in the ancestry of *Brassica* and the tribe Brassiceae which needs further investigations (Lukens et al. 2004).

Like other crop plants, the genome of *Glycine max* L. Merr. (n = 20) has also evolved from an ancestral genome karyotype (n = 11) that lost one chromosome to make n = 10 followed by diploidization (n = 20, Singh and Hymowitz 1988). The soybean genome might have gone through two or more duplications (Blanc and Wolfe 2004; Van et al. 2008) followed by genomic rearrangements (Yan et al. 2003). The whole genome sequence of soybean is underway, and will help in establishing the relationship with legumes and members of other plant families to understand the evolutionary history comprehensively (Schlueter et al. 2008).

In Solanaceae, potato and tomato genomes can be differentiated by few paracentric inversions. Moreover, evidence of translocations and inversions have also been reported in tomato, eggplant and pepper (Zygier et al. 2005).

2.7 Concluding Remarks/Perspectives

About two decades has been invested in comparative genomic research to unravel the consequences of evolutionary forces on shaping the present day genomes of different plant taxa. Following the sequencing of the model plant *Arabidopsis*, the current episode of genome sequencing efforts focused on important agricultural crops has substantially strengthened the ability to study important genes and or gene families conferring novel variation in a particular species, to understand the impact of evolution on shaping the present day genome organization of different crop plants, to begin to explore roles of regulatory elements on genome functionalities for conferring unique traits, and to develop more unified nomenclature for different features of plant taxa. Sequencing of rice, a model for C3 crops, was the first step in the door followed by sequencing of sorghum, a model for C4 crops, and recently of corn, will clarify the basic set of cereal genes, elucidating genome organization of C3 vs C4 mechanisms, grain development, stress related genes and novel features of each plant species. Comparisons of conserved non-coding sequences (CNSs) among different cereal genomes may begin to clarify their role in diversifications of molecular mechanisms conferring important morphological features in different crop species.

The new spectrum of genomes available for comparative genomics coupled with continually-improving bioinformatics methodologies, have improved understanding of synteny and collinearity among different genomes, which will help to translate new genetic information from model species to under-studied or genetically complex crops. Expedited development of new DNA markers for under-explored taxa will jump-start progress in initiating marker-assisted breeding to meet human needs.

In future, computational approaches will have many additional challenging roles, such as predicting new miRNAs which are non-conserved and/or species-specific. Key to community progress will be that websites and databases be amenable

to expansion, user friendly and interoperable to nurture the ability to ask innovative and integrative questions of the burgeoning genomics resources.

Acknowledgements The authors wish to thank Higher Education Commission, Islamabad, Pakistan for providing funds to one of the authors Dr. Mehboob-ur-Rahman for spending 9 months as a postdoc in Plant Genome Mapping Laboratory, Univ Georgia Athens USA.

References

Abdalla AM, Reddy OUK, El-Zik KM et al. (2001) Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. Theor Appl Genet 102:222–229

Adams KL, Cronn R, Percifield R et al. (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc Natl Acad Sci USA 100(8):4649–4654

Ahmad S, Zhang T, Noor-ul-Islam et al. (2007) Identifying genetic variation in Gossypium based on single nucleotide polymorphism. Pak J Bot 39:1245–1250

Ahn S, Tanksley SD (1993) Comparative linkage maps of rice and maize genomes. Proc Natl Acad Sci 90:7980–7984

Akhunov ED, Akhunova AR, Linkiewicz AM et al. (2003) Synteny perturbations between wheat homoeologous chromosomes caused by locus duplications and deletions correlate with recombination rates along chromosome arms. Proc Natl Acad Sci USA 100:10836–10841

Al-Shehbaz IA (1973) The biosystematics of the genus *Thelypodium* (Cruciferae). Contributions to the Gray Herbarium, Harvard University 204:3–148

Ané J, György BK, Brendan K et al. (2004) *Medicago truncatula DMI1* required for bacterial and fungal symbioses in legumes. Science 303:1364–1367

Bailey CD, Koch MA, Mayer M et al. (2006) Toward a global phylogeny of the Brassicaceae. Mol Biol Evol 23(11):2142–2160

Bailey JA, Yavor AM, Massa HF et al. (2001) Segmental duplications: organization and impact within the current human genome project assembly. Genome Res 11:1005–1017

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297

Bates DM (1968) Generic relationships in the Malvaceae, tribe Malveae. Gentes Herbarum 10:117–135

Bates DM, Blanchard OJ (1970) Chromosome numbers in the Malvales. II. New or otherwise noteworthy counts relevant to classification in the Malvaceae, tribe Malveae. Am J Bot 57:927–934

Beasley JO (1940) The production of polyploids in *Gossypium*. J Hered 31:39–48

Beavis B, D.Gessler, Rhee S, Rokhsar D, Doreen M, Lukas M, Huala E, Lincoln S, Lawrence C (2005) Plant biology databases: a needs assessment, http://www.gramene.org/resources/plant_databases.pdf

Bedell JA, Budiman MA, Nunberg A et al. (2005) Sorghum genome sequencing by methylation filtration. PLoS Biol 3(1):e13

Beilstein MA, Al-Shehbaz IA, Kellogg EA (2006) Brassicaceae phylogeny and trichome evolution. Am J Bot 93:607–619

Bennett MD, Leitch IJ (2003) Angiosperm DNA C-values database (release 4_0, Jan. 2003). http://www.rbgkew.org.uk/cval/homepage.html

Bennett MD, Smith JB (1991) Philosophical transactions of the Royal Society B. Biol Sci *B* 334:309–354

Bennetzen JL (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev 15:621–627

Bennetzen J, Ma LJ (2003) The genetic colinearity of rice and other cereals based on genomic sequence analysis. Curr Opin Plant Biol 6:128–133

Bennetzen J, Ma J, Devos KM (2005) Mechanisms of recent genome size variation in flowering plants. Ann Bot 95:127–132

Benton MJ (1993) The Fossil Record 2. Chapman and Hall, New York, USA

Blanc G, Barakat A, Guyot R et al. (2000) Extensive duplication and reshuffling in the *Arabidopsis* genome. Plant Cell 12:1093–1101

Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. The Plant Cell 16:1667–1678

Blanchette M, Green ED, Miller W et al. (2004) Reconstructing large regions of an ancestral mammalian genome in silico. Genome Res 14:2412–2423

Boivin K, Acarkan A, Mbulu R et al. (2004) The Arabidopsis genome sequence as a tool for genome analysis in Brassicaceae. A comparison of the Arabidopsis and Capsella rubella genomes. Plant Physiol 135:735–744

Bonierbale MD, Plaisted RL, Tanksley SD (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120:1095–1103

Bortiri E, Jackson D, Hake S (2006) Advances in maize genomics: the emergence of positional cloning. Curr Opin Plant Biol 9:164–171

Boutin SR, Young ND, Olson T, Yu Z-H, Shoemaker R, Vallejos C (1995) Genome conservation among three legume genera detected with DNA markers. Genome 38:928–937

Bowers JE, Abbey C, Anderson S et al. (2003a) A high-density genetic recombination map of sequence-tagged sites for *Sorghum*, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. Genetics 165:367–386

Bowers JE, Arias MA, Asher R et al. (2005) Comparative physical mapping links conservation of microsynteny to chromosome structure and recombination in grasses. Proc Natl Acad Sci USA 102:13206–13211

Bowers JE, Chapman BA, Rong J et al. (2003b) Unraveling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422:433–438

Brown MS (1980) Identification of the chromosomes of *Gossypium hirsutum* L. by means of translocations. J Hered 71:266–274

Brubaker CL, Paterson AH, Wendel JF (1999) Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome 42:184–203

Buckler ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics: Genome studies and plant molecular genetics. Curr Opin Plant Biol 5:107–111

Buell CR, Yuan Q, Ouyang S et al. (2005) Sequence, annotation, and analysis of synteny between rice chromosome 3 and diverged grass species. Genome Res 15:1284–1291

Burke JM, Tang S, Knapp SJ et al. (2002) Genetic analysis of sunflower domestication. Genetics 161:1257–1267

Burow MD, Simpson CE, Starr JL et al. (2001) Transmission genetics of chromatin from a synthetic amphidiploid to cultivated peanut (*Arachis hypogaea* L.): broadening the gene pool of a monophyletic polyploid species. Genetics 159:823–837

Bustamante CD, Fledel-Alon A, Williamson S et al. (2005) Natural selection on protein-coding genes in the human genome. Nature 437:1153–1157

Cai HW, Morishima H (2002) QTL clusters reflect character associations in wild and cultivated rice. Theor Appl Genet 14:1217–1228

Caicedo AL, Purugganan MD (2005) Comparative plant genomics. Frontiers and prospects. Plant Physiol 138(2):545–547

Canaran P, Buckler ES, Glaubitz JC, Stein L, Sun Q, Zhao W, Ware D (2008) Panzea: an update on new content and features. Nucleic Acids Res 36:1041–1043

Cannon SB, Sterck L, Rombauts S et al. (2006) Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. Proc Natl Acad Sci USA 103:14959–14964

Casasoli M, Derory J, Morera-Dutrey C et al. (2006) Comparison of quantitative trait loci for adaptive traits between oak and chestnut based on an expressed sequence tag consensus map. Genetics 172:533–546

Chao S, Sharp PJ, Worland AJ et al. (1989) RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. Theor Appl Genet 78:495–504

Chapman BA, Bowers JE, Feltus FA et al. (2006) Buffering of crucial functions by paleologous duplicated genes may contribute cyclicality to angiosperm genome duplication. Proc Natl Acad Sci USA 103:2730–2735

Chardon F, Virlon B, Moreau L et al. (2004) Genetic architecture of flowering time in maize as inferred from QTL meta-analysis and synteny conservation with the rice genome. Genetics 162:2169–2185

Chen M, SanMiguel P, Bennetzen JL (1998) Sequence organization and conservation in *sh2/a1*-homologous regions of sorghum and rice. Genetics 148:435–443

Chen ZJ (2007) Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Ann Rev Plant Biol 58:377–406

Chen ZJ, Scheffler BE, Dennis E et al. (2007) Towards sequencing cotton (*Gossypium*) genomes. Plant Physiol 145:1251–1263

Childs KL, Hamilton JP, Zhu W et al. (2007) The TIGR plant transcript assemblies database. Nucleic Acids Res 35:846–851

Chittenden LM, Schertz KF, LIN YR et al. (1994) A detailed RFLP map of *Sorghum bicolor* x *S. propinquum*, suitable for high-density mapping, suggests ancestral duplication of *Sorghum* chromosomes or chromosomal segments. Theor Appl Genet 87:925–933

Choi HK, Luckow MA, Doyle J et al. (2006) Development of nuclear gene-derived molecular markers linked to legume genetic maps. Mol Genet Genomics 276:56–70

Choi HK, Mun JH, Dong-Jin K et al. (2004) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci USA 101(43):15289–15294

Chyi YS, Honeck ME, Sernyk JL (1992) A genetic linkage map of restriction fragment length polymorphism loci for *Brassica rapa* (syn. *campestris*). Genome 35:746–757

Coghlan A, Eichler EE, Oliver SG et al. (2005) Chromosome evolution in eukaryotes: a multi-kingdom perspective. Trends Genet 21:673–682

Conner JA, Conner P, Nasrallah ME et al. (1998) Comparative mapping of the *Brassica S* locus region and its homeolog in *Arabidopsis*: implications for the evolution of mating systems in the Brassicaceae. Plant Cell 10:801–812

Cordeiro GM, Casu R, McIntyre CL et al. (2001) Microsatellite markers from sugarcane (Saccharum spp.) ESTs cross transferable to *erianthus* and sorghum. Plant Sci 160:1115–1123

Cronn RC, Small RL, Haselkorn T et al. (2002) Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. Am J Bot 89:707–725

De Jong JH (1998) High resolution FISH reveals the molecular and chromosomal organization of repetitive sequences in tomato. Cytogenet Cell Genet 81:104

Deutsch M, Long M (1999) Intron-exon structure of eukaryotic model organisms. Nucleic Acids Res 27:3219–3228

Devos KM (2005) Updating the “Crop Circle.” Curr Opin Plant Biol 8:155–162

Devos KM, Gale MD (2000) Genome relationships: the grass model in current research. Plant Cell 12:637–646

Devos KM, Dubcovsky J, Dvorak J et al. (1995) Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination. Theor Appl Genet 91:282–288

Dillon SD, Shapter FM, Henry RJ (2007) Domestication to crop improvement: genetic resources for *Sorghum* and S*accharum* (Andropogoneae). Ann Bot 2007 100:975–989

Ding L, Sabo IA, Berkowicz N et al. (2004) EAnnot: A genome annotation tool using experimental evidence. Genome Res 14:2503–2509

Dixon RA, Sumner LW (2003) Legume natural products. Understanding and manipulating complex pathways for human and animal health. Plant Physiol 131:878–885

Doebley J (2006) Unfallen grains: how ancient farmers turned weeds into crops. Science 312: 1318–1319

Doganlar S, Frary A, Daunay MC et al. (2002a). A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the Solanaceae. Genetics 161:1697–1711

Doganlar S, Frary A, Daunay MC et al. (2002b) Conservation of gene function in the Solanaceae as revealed by comparative mapping of domestication traits in eggplant. Genetics 161:1713–1726
Doust AN, Devos KM, Gadberry MD et al. (2004) Genetic control of branching in foxtail millet. Proc Natl Acad Sci USA 15:9045–9050
Dubcovsky J, Lijavetzky D, Appendino L et al. (1998) Comparative RFLP mapping of Triticum monococcum genes controlling vernalization requirement. Theor Appl Genet 1998:968–975
Duvick J, Fu A, Muppirala U et al. (2008) PlantGDB: a resource for comparative plant genomics. Nucleic Acids Res 36:959–965
Dvorak J, Yang ZL, You FM et al. (2004) Deletion polymorphism in wheat chromosome regions with contrasting recombination rates. Genetics 168:1665–1675
Edwards GA, Mirza MA (1979) Genomes of the Australian wild species of cotton. II. The designation of a new G genome for *Gossypium bickii*. Can J Genet Cytol 21:367–372
Eichler EE, Sankoff D (2003) Structural dynamics of eukaryotic chromosomal evolution. Science 301:793–797
Eisen JA (1998) Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. Genome Res 8:163–167
El Kharbotly A, Leonards-Schippers C, Huigen DJ et al. (1994) Segregation analysis and RFLP mapping of the R1 and R3 alleles conferring race-specific resistance to Phytophthora infestans in progeny of dihaploid potato parents. Mol Gen Genet 242:749–754
El Kharbotly A, Palomino Sanchez C, Salamini F et al. (1996) R6 and R7 alleles of potato conferring race-specific resistance to Phytophthora infestans (Mont.) de Bary identified genetic loci clustering with the R3 locus on chromosome XI. Theor Appl Genet 92:880–884
Endre G, Kereszt A, Kevei Z, Mihacea S et al. GB (2002) A receptor kinase gene regulating symbiotic nodule development. Nature 417:962–966
Endrizzi JE, Turcotte EL, Kohel RJ (1985) Genetics, cytology, and evolution of *Gossypium*. Adv Genetics 23:271–375
Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res 30:1575–1584
Ermolaeva MD, Wu M, Eisen JA et al. (2003) The age of *Arabidopsis thaliana* genome duplication. Plant Mol Biol 51:859–866
Fei Z, Tang X, Alba RM et al. (2004) Comprehensive EST analysis of tomato and comparative genomics of fruit ripening. Plant J 40:47–59
Feldman M, Lupton FGH, Miller TE (1995) Wheats. In: Smartt J, Simmonds NW (eds) Evolution of Crops, 2nd edn. Longman Scientific, London
Feltus FA, Singh HP, Lohithaswa HC et al. (2006) A comparative genomics strategy for targeted discovery of single-nucleotide polymorphisms and conserved-noncoding sequences in orphan crops. Plant Physiol 140:1183–1191
Feuillet C, Keller B (2002) Comparative genomics in the grass family: Molecular characterization of grass genome structure and evolution. Ann Bot (Lond.) 89:3–10
Fernie AR, Willmitzer L (2001) Molecular and biochemical triggers of potato tuber development. Plant Physiol 127:1459–1465
Ferreira ME, Williams PH, Osborn TC (1994) RFLP mapping of *Brassica napus* using doubled haploid lines. Theor Appl Gene 89:615–621
Feuillet C, Keller B (1999) High gene density is conserved at syntenic loci of small and large grass genomes. Proc Natl Acad Sci USA 96:8665–8670
Figdore SS, Ferreira ME, Slocum MK et al. (1993) Association of RFLP markers with trait loci affecting clubroot resistance and morphological characters in *Brassica oleracea* L. Euphytica 69:33–44
Flavell RB, Bennett MD, Smith JB et al. (1974) Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochem Genetic 12:257–269
Foucher F, Morin J, Courtiade J et al. (2003) *DETERMINATE* and *LATE FLOWERING* are two *TERMINAL FLOWER1/CENTRORADIALIS* homologs that control distinct phases of flowering initiation and development in pea. Plant Cell 15:2742–2754
Fransz P, Soppe W, Schubert I (2003) Heterochromatin in interphase nuclei of *Arabidopsis thaliana*. Chromosome Res 11:227–240

Frary A, Nesbitt TC, Frary A et al. (2000) *fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. Science 289:85–88

Frary C, Doganlar S, Frampton A et al. (2003) Fine mapping of quantitative trait loci for improved fruit characteristics from *Lycopersicon Chmielewskii Chromosome* I. Genome 46:235–243

Fryxell PA (1968) A redefinition of the tribe Gossypieae. Bot Gaz 129:296–308

Fryxell PA (1979) The natural history of the cotton tribe. Texas A&M University Press, College Station, Texas

Fryxell PA (1988) Malvaceae of Mexico. Syst Bot Monographs 25:1–522

Fryxell PA (1992) A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). Rheedea 2:108–165

Fulton TM, Van der Hoeven R, Eannetta NT et al. (2002) Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14:1457–1467

Gale M, Devos KM (1998) Plant comparative genetics after 10 years. Science 282:656–659

Gaut BS (2002) Evolutionary dynamics of grass genomes. New Phytol 154:15–28

Gebhardt C, Walkemeier B, Henselewski H et al. (2003) Comparative mapping between potato (*Solanum tuberosum*) and *Arabidopsis thaliana* reveals structurally conserved domains and ancient duplications in the potato genome. Plant J 34:529–541

Gephardt C, Ritter E, Barone A et al. (1991) RFLP maps of potato and their alignment with the homologous tomato genome. Theor Appl Genet 83:49–57

Gingle AR, Yang H, Chee PW et al. (2006) An integrated Web resource for cotton. Crop Sci 46:1998–2007

Giovannoni JJ (2004) Genetic regulation of fruit development and ripening. Plant Cell (Suppl) 16:170–180

Goff SA, Ricke D, Lan TH et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92–100

Graham PH, Vance CP (2003) Legumes: importance and constraints to greater use. Plant Physiol 131:872–877

Grant D, Cregan P, Shoemaker RC (2000) Genome organization in dicots:genome duplication in Arabidopsis and synteny between soybean and Arabidopsis. Proc Natl Acad Sci USA 97:4168–4173

Griffiths-Jones S (2004) The microRNA Registry. Nucleic Acids Res 32:D109–D111

Griffiths-Jones S, Grocock RJ, van Dongen S et al. (2006) miRBase:microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 34:D140–D144

Grover CE, Kim H, Wing RA et al. (2004) Incongruent patterns of local and global genome size evolution in cotton. Genome Res 14:1474–1482

Grube RC, Radwanski ER, Jahn M (2000) Comparative genetics of disease resistance within the solanaceae. Genetics 155:873–887

Gu YQ, Salse J, Coleman-Derr D et al. (2006) Types and rates of sequence evolution at the high-molecularweight glutenin locus in hexaploid wheat and its ancestral genomes. Genetics 174:1493–1504

Guyot R, Yahiaoui N, Feuillet C et al. (2004) In silico comparative analysis reveals a mosaic conservation of genes within a novel colinear region in wheat chromosome 1AS and rice chromosome 5S. Funct Integr Genomics 4:47–58

Hanson PM, Bernacchi D, Green S et al. (2000) Mapping a wild tomato introgression associated with tomato yellow leaf curl virus resistance in a cultivated tomato line. J Am Soc Hort Sci 125:15–20

Hansson B, Kawabe A, Preuss S et al. (2006) Comparative gene mapping in Arabidopsis lyrata chromosomes 1 and 2 and the corresponding *A. thaliana* chromosome 1: recombination rates, rearrangements and centromere location. Genet Res 87:75–85

Hawkins JS, Kim H, Nason JD et al. (2006) Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. Genome Res 16:1252–1261

Henderson IR, Zhang XY, Lu C et al. (2006) Dissecting *Arabidopsis* thaliana DICER function in small RNA processing, gene silencing and DNA methylation pattering. Nat Genet 38:721–725

Hohmann U, Graner A, Endo TR (1995) Comparison of wheat physical maps with barley linkage maps for group 7 chromosomes. Theor Appl Genet 91:618–626

Hu FY, Tao DY, Sacks E et al. (2003) Convergent evolution of perenniality in rice and sorghum. Proc Natl Acad Sci 100:4050–4054

Huang S, van der Vossen EAG, Kuang H et al. (2005) Comparative genomics enabled the isolation of the R3a late blight resistance gene in potato. Plant J 42:251–261

Huang S, Vleeshouwers VG, Werij JS et al. (2004) The R3 resistance to Phytophthora infestans in potato is conferred by two closely linked R genes with distinct specificities. Mol Plant Microbe Interact 17:428–435

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800

Isidore E, Scherrer B, Chalhoub B et al. (2005) Ancient haplotypes resulting from extensive molecular rearrangements in the wheat A genome have been maintained in species of three different ploidy levels. Genome Res 15:526–536

Itoh T, Tanaka T, Barrero RA et al. (2007) Curated genome annotation of *Oryza sativa* ssp. *japonica* and comparative genome analysis with *Arabidopsis thaliana*. Genome Res 17:175–183

Ito Y, Arikawa K, Baltazar A et al. (2005) Rice Annotation Database (RAD): a contig-oriented database for map-based rice genomics. Nucleic Acid Res 33:651–655

Jackson SA, Rokhsar D, Stacey G et al. (2006) Toward a reference sequence of the soybean genome: a multiagency effort. Crop Sci 46:S-55–S-61

Jaillon O, Aury JM, Noel B et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467

Jaiswal P, Ni J, Yap I et al. (2006) Gramene: a bird's eye view of cereal genomes. Nucleic Acids Res 34:717–723

Jannoo N, Grivet L, Chantret N et al. (2007) Orthologous comparison in a gene-rich region among grasses reveals stability in the sugarcane polyploid genome. Plant J 50:574–585

Jiang CX, Wright RJ, El-Zik KM et al. (1998) Polyploid formation created unique avenues for response to selection in Gossypium (cotton). Proc Natl Acad Sci 95:4419–4424

Jones JDG, Dangl JL (2006) The plant immune system. Nature 444:323–329

Judziewicz EJ, Clark LG, Londoño X et al. (1999) American bamboos. Smithsonian Institution Press, Washington, DC, USA

Juretic N, Bureau TE, Bruskiewich RM (2004) Transposable element annotation of the rice genome. Bioinformatics 20:155–60

Karp A. (ed.) (1998) Molecular tools for screening biodiversity (Plants and animals), 1st edn. Chapmann and Hall, London, UK

Kashkush K, Feldman M, Levy AA (2003) Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat Genet* 33:102–106

Kawabe A, Hansson B, Hagenblad J et al. (2006) Centromere locations and associated chromosome rearrangements in *Arabidopsis lyrata* and *A. thaliana*. Genetics 173:1613–1619

Kellogg EA (1998) Relationships of cereal crops and other grasses. Proc Natl Acad Sci USA 97:9121–9126

Kellogg EA, Bennetzen JL (2004) The evolution of nuclear genome structure in seed plants. Am J Bot 91(10):1709–1725

Kennard W, Phillips R, Porter R et al. (1999) A comparative map of wild rice (*Zizania palustris* L. $2n = 2x = 30$). Theor Appl Genet 99:793–799

Kianian SF, Quiros CF (1992) Generation of a *Brassica oleracea* composite RFLP map: linkage arrangements among various populations and evolutionary implications. Theor Appl Genet 84:544–554

Kishimoto N, Higo H, Abe K et al. (1994) Identification of the duplicated segments in rice chromosomes 1 and 5 by linkage analysis of cDNA markers of known functions. Theor Appl Genet 88:722–726

Koch M, Al-Shehbaz IA, Mummenhoff K (2003) Molecular systematics, evolution, and population biology in the mustard family (Brassicaceae). Ann Mo Bot Gard 90:151–171

Koch MA, Kiefer M (2005) Genome evolution among cruciferous plants: a lecture from the comparison of the genetic maps of three diploid species – *Capsella rubella*, *Arabidopsis lyrata* subsp *Petraea*, and *A. thaliana*. Am J Bot 92:761–767

Koch MA, Dobes C, Kiefer C et al. (2007) Supernetwork identifies multiple events of plastid trnF (GAA) pseudogene evolution in the Brassicaceae. Mol Biol Evol 24:63–73
Komatsuda T, Pourkheirandish M, He C et al. (2007) Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. Proc Natl Acad Sci USA 104:1424–1429
Kowalski SP, Lan TH, Feldmann KA et al. (1994) Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organization. Genetics 138:499–510
Kresovich S, Barbazuk B, Bedell JA et al. (2005) Toward sequencing the sorghum genome. A U.S. National Science Foundation-Sponsored Workshop Report. Plant Physiol 138:1898–1902
Krusell L, Madsen LH, Sato S et al. (2002) Shoot control of root development and nodulation is mediated by a receptor-like kinase. Nature 420:422–426
Ku HM, Vision T, Liu J et al. (2000) Comparing sequenced segments of the tomato and Arabidopsis genomes: Large-scale duplication followed by selective gene loss creates a network of synteny. Proc Natl Acad Sci USA 97:9121–9126
Kuittinen H, de Haan AA, Vog C et al. (2004) Comparing the linkage maps of the close relatives *Arabidopsis lyrata* and *A. thaliana*. Genetics 168:1575–1584
Kulikova O, Gualtieri G, Geurts R et al. (2001) Integration of the FISH pachytene and genetic maps of *Medicago truncatula*. Plant J 27:49–58
Kurata N, Yamazaki Y (2006) Oryzabase. An integrated biological and genome information database for rice. Plant Physiol 140:12–17
Lagercrantz U (1998) Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that Brassica genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. Genetics 150:1217–1228
Lagercrantz U, Lydiate D (1996) Comparative genome mapping in *Brassica*. Genetics 144:1903–1910
Lagercrantz U, Putterill J, Coupland G et al. (1996) Comparative mapping in *Arabidopsis* and *Brassica*, fine scale genome collinearity and congruence of genes controlling flowering time. Plant J 9:13–20
Lan TH, DelMonte TA, Reischmann KP et al. (2000) An EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*. Genome Res 10:776–788
Landry BS, Hubert N, Etoh T et al. (1991) A genetic map for *Brassica napus* based on restriction fragment length polymorphisms detected with expressed DNA sequences. Genome 34:543–552
Lawrence CJ, Mary L, Schaeffer TE et al. (2007) MaizeGDB's new data types, resources and activities. Nucleic Acids Res 35:895–900
Lee JM, Grant D, Vallejos CE, Shoemaker RC (2001) Genome organization in dicots. II. Arabidopsis as a 'bridging species' to resolve genome evolution events among legumes. Theor Appl Genet 103:765–773
Lefebvre V, Palloix A, Caranta C et al. (1995) Construction of an intraspecific integrated linkage map of pepper using molecular markers and doubled haploid progenies. Genome 38:112–121
Leister D, Kurth J, Laurie DA et al. (1998) Rapid reorganization of resistance gene homologues in cereal genomes. Proc Natl Acad Sci USA 95:370–375
Levy J, Bres C, Geurts R et al. (2004) A putative Ca^{2+} and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. Science 303:1361–1364
Li C, Zhou A, Sang T (2006) Rice domestication by reducing shattering. Science 311:1936–1939
Liang C, Jaiswa P, Hebbard C et al. (2008) Gramene: a growing plant comparative genomics resource. Nucleic Acids Res 36:947–953
Lin YR, Schertz KF, Paterson AH (1995) Comparative analysis of QTLs affecting plant height and maturity across the poaceae, in reference to an interspecific sorghum population. Genetics 141:391–411
Liu B, Vega JM, Segal G et al. (1998) Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. I. Changes in low copy non-coding DNA sequences. Genome 41:272–277
Liu Z, Moore PH, Ma H et al. (2004) A primitive Y chromosome in papaya marks incipient sex chromosome evolution. Nature 427:348–352
Livingstone KD, Lackney VK, Blauth JR et al. (1999) Genome mapping in Capsicum and the evolution of genome structure in the Solanaceae. Genetics 152:1183–1202

Lukens LN et al. (2004) Genome redundancy and plasticity within ancient and recent Brassica crop species. Biol J Linnean Soc 82:665–674
Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. Genetics 154:459–473
Lysak MA, Koch MA, Pecinka A et al. (2005) Chromosome triplication found across the tribe *Brassiceae*. Genome Res 15:516–525
Lysak MA, Berr A, Pecinka A et al. (2006) Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. Proc Natl Acad Sci USA 103:5224–5229
Lysak MA, Cheung K, Kitschke M et al. (2007) Ancestral chromosomal blocks are triplicated in Brassiceae species with varying chromosome number and genome size. Plant Physiol 145:402–410
Madlung A, Tyagi AP, Watson B et al. (2005) Genomic chnages in synthetic Arabidopsis polyploidy. Plant J 41:221–230
Madsen EB, Madsen LH, Radutoiu S et al. (2003) A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. Nature 425:637–640
Marino CL, Nelson JC, Lu YH et al. (1996) RFLP-based linkage maps of the homoeologous group 6 chromosomes of hexaploid wheat (*Triticum aestivum* L. em. Thell). Genome 39:359–366
Martin DN, Proebsting WM, Hedden P (1997) Mendel's dwarfi ng gene: cDNAs from the *Le* alleles and function of the expressed proteins. Proc Natl Acad Sci USA 94:8907–8911
Matsumoto T, Wu J, Kanamori H et al. (2005) The map based sequence of the rice genome. Nature 436:793–800
Mefford HC, Trask BJ (2002) The complex structure and dynamic evolution of human subtelomeres. Nat Rev Genet 3:91–102
Menacio-Hautea D, Fatokum CA, Kumar L et al. (1993) Comparative genome analysis of mungbean (*Vigna radiata* L. Wilczek) and cowpea (*V. unguiculata*) using RFLP analysis. Theor Appl Genet 86:797–810
Menzel MY, Hasenkampf CA, Stewart JMcD (1982) Incipient genome differentiation in *Gossypium*. III. Comparison of chromosomes of *G. hirsutum* and Asiatic diploids using heterozygous translocations. Genetics 100:89–103
Mickelson-Young L, Endo TR, Gill BS (1995) A cytogenetic ladder-map of the wheat homoeologous group-4 chromosomes. Theor Appl Genet 90:1007–1011
Ming R, Hou S, Feng Y et al. (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). Nature 452:991–997
Ming R, Liu S, Bowers JE et al. (2002) Construction of a *Saccharum* consensus genetic map from two interspecific crosses. Crop Sci 42:570–583
Ming R, Liu S, Lin Y et al. (1998) Alignment of the *Sorghum* and *Saccharum* chromosomes: Comparative genome organization and evolution of a polysomic polyploid genus and its diploid cousin. Genetics 150:1663–1682
Ming R, Moore PH, Wu KK et al. (2005) Sugarcane improvement through breeding and biotechnology. Plant Breeding Rev 27:5–118
Miller W, Makova KD, Nekrutenko A et al. (2004) Comparative genomics. Annu Rev Genomics Hum Genet 5:15–56
Mudge J, Cannon SB, Kalo P et al. (2005) Highly syntenic regions in the genomes of soybean, *Medicago truncatula* and *Arabidopsis thaliana*. BMC Plant Biol 5:15
Mueller LA, Mills AA, Skwarecki B et al. (2008) The SGN comparative map viewer. Bioinformatics 24:422–423
Mueller LA, Solow TH, Taylor N et al. (2005) The SOL genomics network. A comparative resource for Solanaceae biology and beyond. Plant Physiol 138:1310–1317
Mukhtar S, Rahman M, Zafar Y(2002) Assessment of genetic diversity among wheat cultivars using random amplified polymorphic DNA (RAPD) analysis. Euphytica 128:417–425
Muller HJ (1932) Further studies on the nature and causes of gene mutations. Int Congr Genet 61:213–255
Nagaki K, Cheng ZK, Ouyang S et al. (2004) Sequencing of a rice centromere uncovers active genes. Nature Genet 36:138–145

Nagamura Y, Inoque T, Antonio BA et al. (1995) Conservation of duplicated segments between rice chromosomes 11 and 12. Breed Sci 45:373–376

Namuth DM, Lapitan NLV, Gill KS et al. (1994) Comparative RFLP mapping of *Hordeum vulgare* and *Triticum tauschii*. Theor Appl Genet 89:865–872

Naranjo T, Roca A, Goicoechea PG et al. (1987) Arm homoeology of wheat and rye chromosomes. Genome 29:873–882

Nelson JC, Sorrells ME, Van Deynze AE et al. (1995) Molecular mapping of wheat. Major genes and rearrangements in homoeologous groups 4, 5, and 7. Genetics 141:721–731

Nishimura R, Hayashi M, Wu GJ et al. (2002) HAR1 mediates systemic regulation of symbiotic organ development. Nature 420:426–429

Nishiyama T, Fujita T, Shin-I T et al. (2003) Comparative genomics of *Physcomitrella patens* gametophytic transcriptome and *Arabidopsis thaliana*: implication for land plant evolution. Proc Natl Acad Sci USA 100:8007–8012

Ori, N., Eshed, Y., Paran, I., Presting, G., Aviv, D., Tanksley, S., Zamir, D. and Fluhr, R. (1997) The *12C* family from the wilt disease resistance locus *12* belongs to the nucleotide binding, leucine-rich repeat superfamily of plant resistance genes. *Plant Cell*, **9**, 521–532.

Osborn TC (2004) The contribution of polyploidy to variation in *Brassica* species. Physiol Plantarum 121:531–536

Osborn TC, Pires JC, Birchler JA et al. (2003) Understanding mechanisms of novel gene expression in polyploids. Trends Genetic 19:141–147

Ouyang S, Zhu W, Hamilton J et al. (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. Nucleic Acids Res 35:883–887

Oxelman B, Yoshikawa N, McConaughy BL, Luo J, Denton AL, Hall BD (2004) RPB2 gene phylogeny in flowering plants, with particular emphasis on asterids. Mol Phylogenet Evol 32: 462–479

Palmer JD, Soltis DE, Chase MW (2004) The plant tree of life: an overview and some points of view. Am J Bot 91(10):1437–1445

Palmer LE, Rabinowicz PD, O'Shaughnessy AL et al. (2003) Maize genome sequencing by methylation filtration. Science 302:2115–2117

Pan Q, Liu YS, Budai-Hadrian O et al. (2000) Comparative genetics of nucleotide binding site-leucine rich repeat resistance gene homologues in the genomes of two dicotyledons: tomato and Arabidopsis. Genetics 155:309–322

Paterson AH (2002) What has QTL mapping taught us about plant domestication? New Phytol 154:592–608

Paterson AH, Bennetzen JL (2001) Comparative mapping of plant chromosomes. In: Phillips RL, Vasil IK (eds) DNA-based markers in plants, 2nd edn. Kluwer Academic, Dordrecht, pp 101–114

Paterson AH, Bowers JE, Burow MD et al. (2000) Comparative genomics of plant chromosomes. Plant Cell 12:1523–1539

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA 101:9903–9908

Paterson AH, Bowers JE, Peterson DG et al. (2003) Structure and evolution of cereal genomes. Curr Opin Genet Dev 13:644–650

Paterson AH, Chapman BA, Kissinger JC et al. (2006) Many gene and domain families have convergent fates following independent whole-genome duplication events in *Arabidopsis*, *Oryza*, *Saccharomyces* and *Tetraodon*. Trends Genet 22:597–602

Paterson AH, Freeling M, Sasaki T (2005) Grains of knowledge: Genomics of model cereals. Genome Res 15:1643–1650

Paterson AH, Lin YR, Li ZK et al. (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269:1714–1718

Paterson AH, Bowers JE, Bruggmann R et al. (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Patil N, Berno A, Hinds D et al. (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 294:1719–1723

Pedley KF, Martin GB (2003) Molecular basis of Pto-mediated resistance to bacterial speck disease in tomato. Annu Rev Phytopathol 41:215–243

Pereira MG, Lee M (1995) Identification of genomic regions affecting plant height in sorghum and maize. Theor Appl Genet 90:380–388

Periera MG, Lee M, Bramel-Cox P et al. (1994) Construction of an RFLP map in sorghum and comparative mapping in maize. Genome 37:236–243

Petrov DA, Wendel JF (2006) Genome evolution in eukaryotes: The genome size perspective. In: Fox CW, Wolf JB (eds) Evolutionary genetics: Concepts and case studies. Oxford University Press, Oxford, UK

Pfeil BE, Brubaker CL, Craven LA et al. (2002) Phylogeny of Hibiscus and the tribe Hibisceae (Malvaceae) using chloroplast DNA sequences of ndhF and the rpl16 intron. Syst Bot 27:333–350

Pfeil BE, Brubaker CL, Craven LA et al. (2004) Paralogy and orthology in the Malvaceae *rpb2* gene family: investigation of gene duplication in *Hibiscus*. Mol Biol Evol 21(7):1428–1437

Phan HTT, Ellwood SR, Hane JK et al. (2007) Extensive macrosynteny between *Medicago truncatula* and *Lens culinaris* ssp. *Culinaris*. Theor Appl Genet 114:549–558

Phillips LL, Strickland MA (1966) The cytology of a hybrid between *Gossypium hirsutum* and *G. longicalyx*. Can J Genet Cytol 8:91–95

Polhill RM, Raven PH (eds) (1981) Advances in legume systematics, parts 1 and 2. Royal Botanic Gardens, Kew, UK

Prince JP, Pochard E, Tanksley SD (1993) Construction of a molecular linkage map of pepper, and a comparison of synteny with tomato. Genome 36:404–417

Qiu Y, Lee J, Bernasconi-Quadroni F et al. (1999) The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. Nature 402:404–407

Raes J, Vandepoele K, Simillion C et al. (2003) Investigating ancient duplication events in the *Arabidopsis* genome. J Struct Funct Genomics 3:117–129

Rahman M, Hussain D, Zafar Y (2002) Estimation of genetic divergence among elite cotton (*Gossypium hirsutum* L.) cultivars/genotypes by DNA fingerprinting technology. Crop Sci 42:2137–2144

Rahman M, Malik TA, Hussain D et al. (2005) Genetics of resistance to cotton leaf curl virus disease in *Gossypium hirsutum*. Plant Pathol 54:764–772

Rahman M, Yasmin T, Tabassum N et al. (2008) Studying the extent of genetic diversity among *Gossypium arboreum* L. genotypes/cultivars using DNA fingerprinting. Genetic Resour Crop Evol 55:331–339

Rajagopalan R, Vaucheret H, Trejo J et al. (2006) A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. Genes Dev 20:3407–3425

Rhee SY (2005) Bioinformatics. Current limitations and insights for the future. Plant Phy 138:569–570

Reinisch AJ, Dong JM, Brubaker CL et al. (1994) A detailed RFLP map of cotton, *Gossypium hirsutum* × *Gossypium barbadense*: Chromosome organization and evolution in a disomic polyploid genome. Genetics 138:829–847

Rieseberg LH (2001) Chromosomal rearrangements and speciation. Trends Ecol Evol 16:351–358

Rokas A, Williams BL, KING N et al. (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425:798–804

Rong J, Abbey C, Bowers JE et al. (2004) A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). Genetics 166:389–417

Rong J, Bowers JE, Schulze SR et al. (2005) Comparative genomics of *Gossypium* and *Arabidopsis*: Unraveling the consequences of both ancient and recent polyploidy. Genome Res 15:1198–1210

Rong J, Feltus FA, Waghmare VN et al. (2007) Meta-analysis of polyploid cotton QTLs shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. Genetics 176(4):2577–2588

Ross-Ibarra J, Morrell PL, Gaut BS (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. Proc Natl Acad Sci USA 104:8641–8648

Sadowski J, Quiros CF (1998) Organization of an *Arabidopsis thaliana* gene cluster on chromosome 4 including the *RPS2* gene in the *Brassica nigra* genome. Theor Appl Genet 96:468–474
Sadowski J, Gaubier P, Delseny M et al. (1996) Genetic and physical mapping in *Brassica* diploid species of a gene cluster defined in *Arabidopsis thaliana*. Mol Gen Genet 251:298–306
Sakata K, Antonio BA, Mukai Y et al. (2000) INE: a rice genome database with an integrated map view. Nucleic Acids Res 28:97–101
Sakata K, Nagamura Y, Numa H et al. (2002) RiceGAAS: an automated annotation system and database for rice genome sequence. Nucleic Acids Res 30:98–102
Salse J, Bolot S, Throude M et al. (2008) Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. Plant Cell 20:11–24
Salse J, Piegu B, Cooke R et al. (2002) Synteny between Arabidopsis thaliana and rice at the genome level: A tool to identify conservation in the ongoing rice genome sequencing project. Nucleic Acids Res 30:2316–2328
Salse J, Piegu B, Cooke R et al. (2004) New in silico insight into the synteny between rice (Oryza sativa L.) and maize (*Zea mays* L.) highlights reshuffling and identifies new duplications in the rice genome. Plant J 38:396–409
Salvi S, Sponza G, Morgante M et al. (2007) Conserved non-coding genomic sequences controlling flowering time differences in maize. Proc Natl Acad Sci USA 104:11376–11381
Salzberg SL, White O, Peterson J et al. (2001) Microbial genes in the human genome: lateral transfer or gene loss? Science 292:1903–1906
Sato S, Nakamura Y, Asamizu E et al. (2007) Genome sequencing and genome resources in model legumes. Plant Physiol 144:806–820
Schauser L, Roussis A, Stiller J et al. (1999) A plant regulator controlling development of symbiotic root nodules. Nature 402:191–195
Schlueter JA, Scheffler BE, Jackson S et al. (2008) Fractionation of synteny in a genomic region containing tandemly duplicated genes across *Glycine max*, *Medicago truncatula*, and *Arabidopsis thaliana*. J Hered 99:390–395
Schneider K, Weisshaar B, Borchardt DC et al. (2001) SNP frequency and allelic haplotype structure of *Beta vulgaris* expressed genes. Mol Breed 8:63–74
Schranz ME, Lysak MA, Mitchell-Olds T (2006) The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. Trends Plant Sci 11:535–542
Schranz ME, Windsor AJ, Song B et al. (2007) Comparative genetic mapping in *Boechera stricta*, a close relative of Arabidopsis. Plant Physiol 144:286–298
See DR, Brooks S, Nelson JC et al. (2006) Gene evolution at the ends of wheat chromosomes. Proc Natl Acad Sci 103:4162–4167
Seijo G, Lavia GI, Fernandez A et al. (2007) Genomic relationships between the cultivated peanut (*Arachis hypogaea*, Leguminosae) and its close relatives revealed by double GISH. Am J Bot 94:1963–1971
Shaheen T, Rahman M, Zafar Y (2006) Chloroplast RPS8 gene of cotton reveals the conserved nature through out taxa. Pak J Bot 38:1467–1476
Shaked H, Kashkush K, Ozkan H et al. (2001) Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. Plant Cell 13:1749–1759
Shinozaki K, Ohme M, Tanaka M et al. (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5:2043–2049
Shultz JL, Kurunam D, Shopinski K et al. (2006) The Soybean Genome Database (SoyGD): a browser for display of duplicated, polyploid, regions and sequence tagged sites on the integrated physical and genetic maps of Glycine max. Nucleic Acids Res 34:758–765
Simons G, Groenendijk J, Wijbrandi J et al. (1998) Dissection of the Fusarium *I2* gene cluster in tomato reveals six homologs and one active gene copy. Plant Cell 10:1055–1068
Singh R, Hymowitz T (1988) The genomic relationship between *Glycine max* and *G. soja* Sieb., and Zucc. as revealed by pachytene chromosome analysis. Theor Appl Genet 76: 705–711
Singh NK, Raghuvanshi S, Srivastava SK et al. (2004) Sequence analysis of the long arm of rice chromosome 11 for rice-wheat synteny. Funct Integ Genomics 4:102–117

Singh NK, Dalal V, Batra K et al. (2007) Single-copy genes define a conserved order between rice and wheat for understanding differences caused by duplication, deletion, and transposition of genes. Funct Integr Genomics 7:17–35

Slade AJ, Fuerstenberg S, Loeffler D et al. (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nat Biotechnol 23:75–81

Slocum MK, Figdore SS, Kennard WC et al. (1990) Linkage arrangement of restriction fragment lenght polymorphism loci in *Brassica oleracea*. Theor Appl Genet 80:57–64

Soltis DE, Soltis PS, Tate JA (2004) Advances in the study of polyploidy since *Plant Speciation*. New Phytol 161:173–191

Song K, Lu P, Tang K et al. (1995) Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. Proc Natl Acad Sci USA 92:7719–7723

Song KM, Suzuki JY, Slocum MK et al. (1991) A linkage map of *Brassica rapa* (syn. *campestris*) based on restriction fragment length polymorphism loci. Theor Appl Genet 82:296–304

Song R, Segal G, Messing J (2004) Expression of the sorghum 10-member kafirin gene cluster in maize endosperm. Nucleic Acids Res 32:e189

Sorrells ME, Rota ML, Catherine E et al. (2003) Comparative DNA sequence analysis of wheat and rice genomes. Genome Res 13:1818–1827

Speulman E, Bouchez D, Holub EB (1998) Disease resistance gene homologs correlate with disease resistance loci of Arabidopsis thaliana. Plant J 14:467–474

Stanhope MJ, Lupas A, Italia MJ et al. (2001) Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. Nature 411:940–944

Stracke S, Kistner C, Yoshida S et al. (2002) A plant receptor-like kinase required for both bacterial and fungal symbioses. Nature 417:959–962

Suwabe K, Tsukazaki H, Iketani H et al. (2006) Simple sequence repeat-based comparative genomics between *Brassica rapa* and *Arabidopsis thaliana*: The genetic origin of clubroot resistance. Genetics 173:309–319

Swigonova Z, Bennetzen JL, Messing J (2005) Structure and evolution of the r/b chromosomal regions in rice, maize and sorghum. Genetics 169:891–906

Swigonova Z, Lai J, Ma J et al. (2004) Close split of sorghum and mazie genome proginators. Genome Res 14:1916–1923

Taher L, Rinner O, Garg S et al. (2004) AGenDA: gene prediction by cross-species sequence comparison. Nucleic Acids Res 32:305–308

Tang B, Bowers JE, Wang X et al. (2008a) Synteny and colinearity in plant genomes. Science 320:486–488

Tang B, Wang X, Bowers JE et al. (2008b) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. Genome Res 18:1944–1954

Tanksley SD (2004) The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. Plant Cell (Suppl) 16:181–189

Tanksley SD, Ganal MW, Prince JP et al. (1992) High density molecular linkage maps of the tomato and potato genomes. Genetics 132:1141–1160

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

The Rice Chromosome 3 Sequencing Consortium (2005) Sequence, annotation, and analysis of synteny between rice Chromosome 3 and direrged grass species. Genome Res 15:1284–1291

Thomas BC, Pedersen B, Freeling M (2006) Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. Genome Res 16:934–946

Thorup TA, Tanyolac B, Livingstone KD et al. (2000) Candidate gene analysis of organ pigmentation loci in the Solanaceae. Proc Natl Acad Sci USA 97:11192–11197

Town CD, Cheung F, Maiti R et al. (2006) Comparative genomics of Brassica oleracea and Arabidopsis thaliana reveal gene loss, fragmentation, and dispersal after polyploidy. Plant Cell 18:1348–1359

Teutenico RA, Osborn TC (1994) Mapping of RFLP and qualitative trait loci in *Brassica rapa*, and comparison to linkage maps of *B. napus*, *B. oleracea*, and *Arabidopsis thaliana.* Theor Appl Genet 89:885–894

Tranquilli G, Dubcovsky J (2000) Epistatic interaction between vernalization genes *Vrn-Am1* and *Vrn-Am2* in diploid wheat. J Hered 91:304–306

Truco MJ, Hu J, Sadowski J et al. (1996) Inter- and intra-genomic homology of the Brassica genomes: implications for their origin and evolution. Theor Appl Genet 93:1225–1233

Tsudzuki J, Nakashima K, Tsudzuki T et al. (1992) Chloroplast DNA of black pine retains a residual inverted repeat lacking rRNA genes: Nucleotide sequence of *trnQ*, *trnK*, *psbA*, *trnI* and *trnH* and the absence of *rps16*. Mol Gen Genet 232:206–214

Tuskan GA, Difazio S, Jansson S et al. (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science 313:1596–1604

Udall JA, Quijada PA, Lambert B et al. (2006) Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (*Brassica napus* L.). 2. Identification of alleles from unadapted germplasm. Theor Appl Genet 113:597–609

Ulloa M, Saha S, Jenkins JN et al. (2005) Chromosomal assignment of RFLP linkage groups harboring important QTLs on an intraspecific cotton (*Gossypium hirsutum* L.) Joinmap. J Hered 96:132–144

Uzunova M, Ecke W, Weissleder K et al. (1995) Mapping the genome of rapeseed (*Brassica napus* L.) I. Construction of an RFLP linkage map and localization of QTLs for seed glucosinate content. Theor Appl Genet 90:194–204

Van K, Kim DH, Cai CM et al. (2008) Sequence level analysis of recently duplicated regions in soybean (*Glycine max* (L.) Merr.) Genome DNA Res 15:93–102

Van der KE, Tanksley S (2003) The making of a bell pepper-shaped tomato fruit: identification of loci controlling fruit morphology in Yellow Stuffer tomato. Theor Appl Genet 107: 139–147

Van Deynze AE, Dubcovsky J, Gill KS et al. (1995) Molecular-genetic maps for chromosome 1 in *Triticeae* species and their relation to chromosomes in rice and oats. Genome 38:47–59

Vandepoele K, Simillion C, Van de Peer Y (2003) Evidence that rice and other cereals are ancient aneuploids. Plant Cell 15:2192–2202

Vettore AL, da Silva FR, Kemper EL et al. (2003) Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. Genome Res 13:2725–2735

Vincentz M, Cara FAA, Okura VK et al. (2004) Evaluation of monocot and eudicot divergence using the sugarcane transcriptome. Plant Physiol 134:951–959

Vinogradov AE (1999) Intron-genome size relationship on a large evolutionary scale. J Mol Evol 49:376–384

Vision TJ, Brown DG, Tanksley SD (2000) The origins of genome duplications in *Arabidopsis*. Science 290:2114–2117

Wall PK, Leebens-Mack J, Muüller KF et al. (2008) PlantTribes: a gene and gene family resource for comparative genomics in plants. Nucleic Acids Res 36:970–976

Wang B, Brendel V (2006) Genomewide comparative analysis of alternative splicing in plants Proc Natl Acad Sci 103:7175–7180

Wang J, Tian L, Lee H-S et al. (2006) Genome wide non-additive gene regulation in Arabidopsis allotetraploids. Genetics 172:507–517

Wang WY, Barratt BJ, Clayton DG et al. (2005) Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet 6:109–118

Ware DH, Jaiswal P, Ni J et al. (2002) Gramene, a tool for grass genomics. Plant Physiol 130:1606–1613

Warwick SI, Al-Shehbaz IA (2006) Brassicaceae: chromosome number index and database on CD-Rom. Plant Syst Evol 259:237–248

Warwick SI, Al-Shehbaz IA, Sauder CA (2006) Phylogenetic position of *Arabis arenicola* and generic limits of *Eutrema* and *Aphragmus* (Brassicaceae) based on sequences of nuclear ribosomal DNA. Can J Bot 84:269–281

Weeden N (2007) Genetic changes accompanying the domestication of *Pisum sativum*: Is there a common genetic basis to the 'domestication syndrome' for legumes? Ann Bot 100: 1017–1025

Weeden NL, Muehlbauer FJ, Ladizinsky G (1992) Extensive conservation of linkage relationships between pea and lentil genetic maps. J Hered 83:123–129

Wei F, Coe E, Nelson W et al. (2007) Physical and genetic structure of the maize genome reflects its complex evolutionary history. PLoS Genet 3:e123

Wendel JF (1989) New World tetraploid cottons contain Old World cytoplasm. Proc Natl Acad Sci USA 86:4132–4136

Wendel JF, Albert VA (1992) Phylogenetics of the cotton genus (*Gossypium*): Character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. Syst Bot 17:115–143

Wendel JF, Cronn RC (2003) Polyploidy and the evolutionary history of cotton. Adv Agron 78:139–186

Wendel JF, Cronn RC, Alvarez I et al. (2002) Intron size and genome size in plants. Mol Biol Evol 19:2346–2352

Whitelaw CA, Barbazuk WB, Pertea G et al. (2003) Enrichment of gene-coding sequences in maize by genome filtration. Science 302:2118–2120

Whitkus R, Doebley J, Lee M (1992) Comparative genetic mapping of sorghum and maize. Genetics 132:1119–1130

Wicker T, Yahiaoui N, Keller B (2007) Contrasting rates of evolution in *Pm3* Loci from three wheat species and rice. Genetics 177:1207–1216

Wicker T, Guyot R, Yahiaoui N et al. (2003) CACTA Transposons in Triticeae. A diverse family of high-copy repetitive elements. Plant Physiol 132:52–63

Wills DM, Burke JM (2007) QTL analysis of the early domestication of sunflower. Genetics 176:2589–2599

Windsor AJ, Mitchell-Olds T (2006) Comparative genomics as a tool for gene discovery. Curr Opin Biotech 17:161–167

Wu F, Mueller LA, Crouzillat D et al. (2006) Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the Euasterid plant clade. Genetics 174:1407–1420

Yahiaoui N, Brunner S, Keller B (2006) Rapid generation of new powdery mildew resistance genes after wheat domestication. Plant J 47:85–98

Yahiaoui N, Srichumpa P, Dudler R et al. (2004) Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene *Pm3b* from hexaploid wheat. Plant J 37:528–538

Yan HH, Mudge J, Kim DJ et al. (2003) Estimates of conserved microsynteny among the genomes of *Glycine max, Medicago truncatula*, and *Arabidopsis thaliana*. Theor Appl Genet 106:1256–1265

Yang YW, Lai KN, Tai PY et al. (1999) Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between Brassica and other angiosperm lineages. J Mol Evol 48:597–604

Yang T-J, Kim JS, Kwon S-J et al. (2006) Sequence-level analysis of the diploidization process in the triplicated FLOWERING LOCUS C region of *Brassica rapa*. Plant Cell 18:1339–1347

Yogeeswaran K, Frary A, York TL et al. (2005) Comparative genome analyses of *Arabidopsis* spp.: Inferring chromosomal rearrangement events in the evolutionary history of *A. thaliana.* Genome Res 15:505–515

Young ND, Cannon SB, Sato S et al. (2005) Sequencing the Gene spaces of Medicago truncatula and Lotus japonicus. Plant Physiol 137:1174–1181

Yu J, Hinze L, Frelichowski J et al. (2008) CottonDB: Cotton Genome Database. Intl Cotton genome Initiative. Plant and Animal Genome Conference (Jan 12–16, 2008) W 253

Yu J, Hu S, Wang J et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. Spp. *indica*). Science 296:79–92

Yu J, Wang J, Lin W et al. (2005) The Genomes of *Oryza sativa*: a history of duplications. PLoS Biol 3:e38

Yuan Q, Ouyang S, Wang A et al. (2005) The Institute for Genomic Research Osa1 rice genome annotation database. Plant Physiol 138:18–26

Zhao XP, Ji YF, Ding XL et al. (1998) Macromolecular organization and genetic mapping of a rapidly evolving chromosome-specific tandem repeat family (B77) in cotton (Gossypium). Plant Mol Biol 38(6):1031–1042

Ziolkowski PA Kaczmarek M, Babula D et al. (2006) Genome evolution in *Arabidopsis*/*Brassica*: conservation and divergence of ancient rearranged segments and their breakpoints. Plant J 47:63–74
Zhu H, Choi HK, Cook DR et al. (2005) Bridging model and crop legumes through comparative genomics. Plant Physiol 137:1189–1196
Zhu W, Buell CR (2007) Improvement of whole-genome annotation of cereals through comparative analyses. Genome Res 17:299–310
Zygier S, Chaim AB, Efrati A et al. (2005) QTLs mapping for fruit size and shape in chromosomes 2 and 4 in pepper and a comparison of the pepper QTL map with that of tomato. Theor Appl Genet 111:437–445

Chapter 3
Functional Genomics For Crop Improvement

Seedhabadee Ganeshan, Pallavi Sharma, and Ravindra N. Chibbar

Abstract Plant breeding has had a tremendous influence on crop improvement. However, due to dwindling germplasm resources, identification of variability for incorporation into new cultivars is becoming more difficult. Therefore, there has been recourse to alternative approaches including mutagenesis, tissue culture and genetic transformation to aid breeding programs. Furthermore, with the vast repertoire of genome-wide data from different expression profiling techniques such as microarrays, more subtle understanding of gene expression is being obtained and is further helping plant breeders to entertain a different selection approach based on expression quantitative traits to maximize combinations of genes capable of conferring high performance. In this chapter, we review some of the aspects of plant breeding and the influence functional genomics has on breeding programs. Some of the challenges to functional genomics and breeding come from establishment of high-throughput transformation systems to assess gene function, which is limiting functional characterization of numerous genes in their respective crops. Therefore, this chapter also focuses on the need to gain better understanding of the development of gene transfer systems for crop plants to make use of the array of available gene information data.

3.1 Introduction

The significant influence plant breeding has on crop improvement has not always been fully appreciated. Crop improvement through plant breeding has been a major factor in meeting the food requirements of the growing world population to date. Among the crop plants, the eminence of cereals in sustaining today's human dietary food supplement is iterated by a staggering 66% contribution to the total food consumption (Borlaug 2002). Of the cereals, wheat, maize and rice form the bulk of the production. The global consumption of each of these three major cereals varies,

S. Ganeshan, P. Sharma, and R.N. Chibbar (✉)
Department of Plant Sciences, College of Agriculture and Bioresources,
University of Saskatchewan, Saskatoon, SK, Canada
e-mail: ravi.chibbar@usask.ca

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_3, © Springer Science+Business Media B.V. 2010

in essence due to historical and cultural influences and environmental sustainability for growth in specific areas. Rice, for example, is the staple food for more than three billion people, mostly from the developing countries. However due to dwindling arable lands and the ever-increasing world population, an increase in productivity solely through conventional breeding programs will not suffice. The trail into this cul-de-sac is further compounded by environmental constraints. Weeds, pathogens and pests account for about 40% loss in total crop production, in addition to the 10–30% post harvest crop losses due to biotic and abiotic factors (Oerke et al. 1994). Furthermore the demand for food by the ever-increasing world population will become more significant. Thus, faced with these immense challenges of meeting the global food requirements, there is a need to revitalize our crop production systems through all possible avenues. Conventional as well as newer technological approaches have become mutually contributing in this regard. The major challenges now lie in the core of fundamental plant functioning, dictated by the genetic complexities. Therefore, dissection of the genomic intricacies will unravel clues governing the cascades of events needed for productivity enhancement per se, as well as for biotic and abiotic challenges that indirectly affect productivity.

Breeding programs have been and will continue to strive towards producing varieties with increased productivity. Newer enabling technologies are likely to be the much-awaited catalysts to enhance crop productivity to meet the growing food demands of an increasing world population. Biotechnological approaches have branched into multi-faceted dimensions over the last few decades and provide sets of tools to complement breeding efforts aimed at crop improvement in various ways. These sets of tools encompass such areas as in vitro culture technology, transgene technology, DNA-based mapping and DNA-based marker-assisted selection, gene cloning and sequencing technologies. Furthermore, tools for analysis of the transcriptome, proteome and metabolome have also attained an unprecedented level of sophistication. These tools have been referred to as being part of the functional genomics toolbox (Colebatch et al. 2002). Although the gamut of implements may be perceived as being task-specific, in effect they complement one another at one point or the other, resulting in a more refined understanding of the genome and identification of its functional attributes, leading to the application of functional genomics tools to plant breeding practices (Fig. 3.1). Appropriate steps can thus be taken for incorporation of desirable genomic strategies into respective plant breeding programs. Transgene technologies can be utilized for studying the expression of genes of interest using mutant lines or for gene-inactivation approaches and over-expression in plants. Although sequencing projects effectively dissect the genomes, many of the sequenced segments have unknown functions. Therefore, availability of transformation systems for studying the expression of identified genes or genes of unknown functions in planta would be of value for functional assessment to maximize performance of breeding lines. Furthermore, the extensive developments in gene expression profiling technologies also augur well for complementing breeding programs, especially with the possibility of selecting for expression traits by using expression quantitative trait loci (eQTL) mapping data. This chapter covers studies related to functional genomics tools and their application for crop improvement.

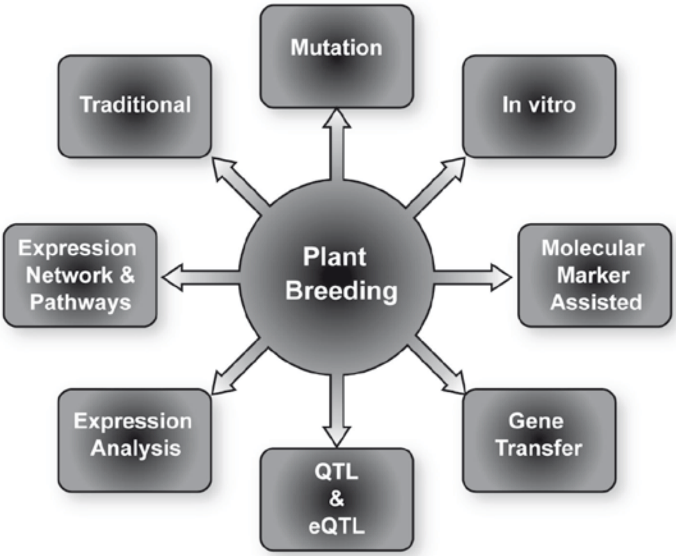


Fig. 3.1 Plant breeding as influenced by various approaches from the classical systems to structural and functional genomics strategies

3.2 Crop Improvement Strategies

Plant breeding is a form of expedited evolution for economically important plants as a means to develop new, improved varieties. It has been described as an art, because there is an observational attribute from the breeders' perspective in singling out plants with desirable trait(s) from a plant population. Therefore, selection of individuals with desirable traits was perceived as the initial beginnings of plant breeding (Sleper and Poehlman 2006). Consequently, the early gatherers would in all likelihood qualify as breeders as they intentionally selected for desirable attributes, such as palatability of plant products and ease of harvest, during the process of domestication of plants and probably did not have much influence on the general structure of crop plants (Hancock 2004). However, before the re-discovery of Mendel's laws of genetics, successful breeders were described as those who were able to carefully observe and diligently identify plant variants with improved qualities for further propagation (Sleper and Poehlman 2006).

In spite of the technological advances in all areas of plant sciences, plant breeding will continue to form the basis of all crop improvement ventures designed to meet the requirements of the world's food supply. Although plant breeding primarily encompasses genetic principles, the inter-dependence of several other disciplines such as botany, agronomy, physiology, biochemistry and more recently biotechnology and molecular biology cannot be iterated enough for further successes in plant breeding. The basic requirements for plant breeding include identification of genetic variability,

followed by combinations of this variability to generate new plant types with desired qualities. However, due to the dwindling genetic diversity, plant breeders have to generate variability for use in their breeding programs. As a first step, traditional breeding utilized desirable germplasm from wild species, land races and distant relatives, or created new variability through chemical and physical mutagenic treatments. The recognition of the phenomenon of somaclonal variation derived from plant tissue cultures (Larkin and Scowcroft 1981) became an added source of variation with lots of promise for plant breeders. Further technological advances of promise to further enhance breeding programs included the advent of molecular markers for expedited selection for traits of interest, leading to marker-assisted selection terminology, and more recently the ability to identify quantitative trait loci (QTLs) with such markers. From the beginning of this decade functional genomics tools have been heralded as yet other potential contributors to plant improvement programs. Therefore, to better appreciate the potential role of functional genomics in crop improvement, an overview of the various developments in breeding practices over the years will be useful.

3.3 Classical Breeding

The classical breeding approaches essentially received a scientific scaffold with the re-discovery of the laws of inheritance at the turn of the last century (Kang et al. 2007). Generally, plant breeding has involved increasing adaptation of cultivars for specific environments to increase productivity. It includes identification of variability in locally available germplasm or from introductions such as land races or other breeding programs, wild species or genera. To compensate for lack of genetic variability, plant breeders resorted to identification of spontaneous mutants or even mutations induced by physical or chemical means. The identified variants from all these sources served as the basis for establishing parental lines for eventual hybridization and selection for desired qualitative or quantitative traits. Depending on the traits to be selected for and whether the plants are self-pollinating or cross-pollinating, the breeder sets realistic objectives to maximize identification of desired traits in the progeny of the crosses. In summary the objectives of a breeding program determine choice of parents, choice of breeding method, selection criteria and selection timing.

Induced mutagenesis was one of the most significant additions to plant breeding methodologies in the early part of the last century. Due to dwindling of naturally occurring variability, plant breeders needed to find other potential sources of variability. Therefore, there was recourse to screening for mutations for incorporation of new traits into plant breeding programs. The occurrence of natural mutations and their potential use in breeding were recognized as early as the 1900s by de Vries (cited in van Harten 1998). With the discovery of the possibility of physically generating mutations by X-rays in *Drosophila* (Muller 1927; Muller 1928) and in barley (Stadler 1928), a whole new field of induced mutagenesis research emerged and

eventually became a contributing field in plant breeding. There are 2,570 mutant varieties listed with the FAO/IAEA Mutant Varieties Database (FAO/IAEA 2006). About 70% of the mutants in the database were direct mutants, implying that a selected mutant was advanced over generations before being released as a new variety, while the remaining mutants were used in crosses with other varieties in breeding programs (Ahloowalia et al. 2004). The first variety to be derived from mutagenic X-ray irradiation was a tobacco cultivar, Chlorina, which was commercially released in 1934 (Tollenaar 1934 cited in van Harten 1998). About two decades later, a commercial variety of mustard, cv. Primex, was released from X-ray-induced mutation treatments. Of the direct mutant varieties released, radiation was used in about 89% of these and 75% of the overall mutants in the database were in crops (Ahloowalia et al. 2004). With the advent of sensitive variant DNA sequence detection techniques, induced mutation has received renewed interest with the mutation strategy termed TILLING (Targeting Induced Local Lesions IN Genomes) (McCallum et al. 2000). TILLING is combination of classical chemical mutagenesis and high-throughput molecular biology techniques for the identification of desired mutants (Chapter 6). It is foreseen that more effective and precise mutation identification using TILLING populations for incorporation into breeding programs will become routine.

3.4 Technology-Based Breeding

Technology driven improvement of plants has become an integral part of breeding programs. The advent of plant tissue culture, followed by the recognition of the occurrence of the phenomenon of somaclonal variation (Larkin and Scowcroft 1981) and subsequently the possibility of using doubled haploids offered immense possibilities in breeding programs. With the advent of molecular markers and the possibility of using such markers for early generation selection for traits of interest, plant breeders received an additional tool. Subsequent expansion of the technology has led to scanning for QTLs and the exploring for genotype × environment interactions in mapping populations in a more expedited manner. Concurrent developments in genetic transformation methodologies from the early 1980s enabled the transfer of exogenous DNA to plants both by physical as well as biological means. Gene transfer to plants nowadays is not only viewed as a way of producing transgenic plants, but also as a method to study functions of cloned genes to eventually enable the breeder to maximize combinations of genes likely to positively impact the performance of the plants under particular circumstances.

The application of plant tissue culture to plant breeding programs became relevant for clonal mass propagation of breeding lines, especially for those long life cycle plants such as trees, generation of pathogen-free stocks for maintenance as well as exchange of such stocks among breeders. The application of tissue culture has also played an important role in the clonal propagation of ornamental plants and is still an important technique in this field (for review see Rout et al. 2006). However, for many years it was believed that large populations of single somatic cells could be

perceived as cells of microorganisms and could be used to specifically search for biochemical mutants (Strauss et al. 1980). The possibility of establishing callus, suspension and protoplast cultures in many plant species augured well for such a screening approach for mutant cell lines. In the early 1980s and onwards there were numerous reports on the exposure of somatic cells to defined stimuli in the culture media and selection for cell lines tolerant or resistant to such stimuli, a practice referred to as "in vitro selection". In vitro selection was quickly embraced by plant pathologists and plant breeders alike to select for resistance to plant pathogens producing toxins involved in pathogenesis (for reviews see Buiatti and Ingram 1991; Svabova and Lebeda 2005). There has also been an interest over the years in selecting for salt tolerance (e.g., Karadimova and Djambova 1993; Winicov 1996; Zhang et al. 2001; Dziadczyk et al. 2003; Queiros et al. 2007; Hamrouni et al. 2008; He et al. 2009) and herbicide resistance (e.g., Kinoshita and Mori 1991; Pofelis et al. 1992; Baillie et al. 1993; Escorial et al. 1996; Taregyan et al. 2001; Yu et al. 2004; Gurel et al. 2008) using such culture systems.

The occurrence of somaclonal variation in tissue cultures was yet another potential source of genetic variability for use by plant breeders in their breeding programs. There was considerable interest in the potential of somaclonal variation, possibly in combination with in vitro selection, for plant improvement. However, there was also an apprehension among many breeders as to the value of somaclonal variation since valuable genetic stocks could have undesirable changes due to in vitro culture, and the availability of tissue culture methodologies for maintenance of genetic stocks in gene banks were almost non-existent (Jain 2001). Nonetheless, many commercial varieties of economically important plants have been released as a result of somaclonal variation in combination with in vitro selection strategies. Somaclonal variation has been used to induce several improved attributes such as disease resistance, salinity and drought tolerance, insect resistance and nutritional quality in diverse plants such as rice, wheat, maize, banana, tomato, sugar cane and flax amongst others (for review see Jain 2001).

One of the most important developments in the use of tissue culture for breeding was the ability to use embryo rescue to regenerate hybrids from crosses showing limited sexual compatibility (Raghavan 1985; Raghavan 1986). Embryo rescue is still being used in breeding programs for the generation of genetic variability and incorporation of desirable traits from inter-specific or inter-generic hybrids (for review see Sharma et al. 1996). Embryo rescue to develop improved germplasm has been successfully used in legumes (Clarke et al. 2006; Fratini and Ruiz 2006; Fratini and Ruiz 2008), cereals (Rodrangboon et al. 2002; Tikhenko et al. 2008), oilseed crops (Luhs and Friedt 1994; Muangprom et al. 2006), horticultural crops (Chen et al. 2004; Drew et al. 2006; Yang et al. 2007; Tian and Wang 2008) and many other economically important plant species. A more sophisticated approach to the creation of genetic variability was through somatic cell hybridization by fusion of protoplasts, and in the late 1970s several significant breakthroughs occurred in addition to publication of a number of reports pertaining to basic protoplast isolation and culture (for reviews see Lynch et al. 1993; Waara and Glimelius 1995; Davey et al. 2005; Liu et al. 2005). Due to the extensive efforts undertaken for several important plants

with regards to protoplast fusion and somatic hybridization, several reviews relating to specific plants such as citrus (Grosser et al. 2000), potato (Millam et al. 1995; Orczyk et al. 2003) and tomato (Lefrancois et al. 1993) have been published. Many of the research efforts focused on the most economically important crop plants such as rice (e.g., Zhu et al. 2004; Feng et al. 2006), wheat (Ge et al. 2006), the Brassicaceae (Wang et al. 2003, 2005, 2006) and potato (Bidani et al. 2007; Thieme et al. 2008).

The other major influence from in vitro culture technology on breeding practices has been the production of doubled haploid (DH) plants (Touraev et al. 2009). Generally in conventional plant breeding, selfing over a number of generations is necessary to achieve homozygozity starting from a heterozygous population (Nei 1963). The utility of a haploid system was recognized in the late 1940s with the identification of maize parthenogenic haploids (Chase 1949; Chase 1952). The first report on the occurrence of natural haploids was in *Datura stramonium* L. (Blakeslee et al. 1922). Subsequently there were several reports on the identification and generation of haploids through X-ray treatment or as a result of induced parthenogenesis or natural occurrence of parthenogenesis after wide crosses. However a major breakthrough in the 1960s was achieved with the culture of *Datura* anthers in vitro with the regeneration of haploid plants (Guha and Maheshwari 1964; Guha and Maheshwari 1966). In addition to spontaneous doubling of chromosomes, the use of colchicine for chromosome doubling in haploids became a useful tool to maximize the number of doubled haploid plants for breeding purposes. Today with more improvements it has been possible to culture microspores as well as ovules for the production of doubled haploids in many species. In the beginning, however, lack of efficient production of doubled haploids from the relatively recalcitrant species was still hindering adoption of this approach for breeding purposes. The observation that in wide crosses of *Hordeum vulgare* × *Hordeum bulbosum* elimination of the latter set of *bulbosum* chromosomes occurred (Kasha and Kao 1970), augured well for producing doubled haploid plants and the barley system has extensively been used in barley breeding programs worldwide. Other examples of wide crosses for doubled haploid production based on chromosome elimination include wheat x maize for production of doubled haploid durum wheat (Sarrafi et al. 1994) and *Solanum tuberosum* × *S. phureja* (Hermsen and Verdenius 1973). With advances in tissue culture, haploid plant can now be regenerated from in vitro cultured microspores with high efficiency in many plant species for incorporation in breeding programs.

3.5 Expression Profiling

Gene expression control is one of the key regulatory mechanisms in living cells for maintenance and implementation of their functions. Such regulation is brought about by the epigenetic and transcriptional regulation of genes leading to synthesis of RNA molecules. Degradation of these molecules is a further regulatory checkpoint in the regulation of gene expression. Thus functioning of the cells in concert with other cells, tissues and organs is coordinately regulated temporally, spatially,

developmentally as well as environmentally. There has been a long-standing interest in the study of gene expression and regulation. One of the earliest methods available for determination of gene expression has been RNA gel blot hybridization. While still in use, it has disadvantages of being labor-intensive, cumbersome and requiring large amounts of RNA. Furthermore, it requires use of radio-labeled probes, although non radio-active methods based on chemi-luminescence are nowadays available. Another slightly more sensitive method is the nuclease protection assay (NPA), which has some of the same disadvantages as RNA gel blot hybridization. However, with NPA there is a possibility of studying multiple RNA targets, provided the lengths of the hybridizing probes are distinctly different. Nonetheless the sensitivity, low throughput nature and non-quantitative or at best semi-quantitative nature of these two methods preclude the requirements for present gene expression analyses aimed at interactions among genes. Although in situ hybridization methods have in recent years become more sensitive, their successful establishment is still determined by user experience, extensive optimization, labor-intensiveness and probably irreproducibility among laboratories. Moreover, due to the feverish pace with which the repertoire of genome sequence data is accumulating, the need for more sensitive and high throughput expression analysis methods became imperative.

The advent of the polymerase chain reaction (PCR) (Mullis and Faloona 1987; Saiki et al. 1988) with unsurpassed sensitivity for amplification of DNA from minute amounts subsequently became an added tool for gene expression analysis. Since the first report of the possibility of reverse-transcribing mRNA for amplification using PCR (Veres et al. 1987), the development of the RTPCR (reverse-transcriptase PCR) approach has had significant influence on gene expression analysis. However, sensitivity and quantification were still debatable. Further refinement of quantification aspects during PCR amplification led to the development of the kinetic or real-time PCR with monitoring of amplification at each cycle and the possibility of relating the amplified products to the starting amount of template (Higuchi et al. 1993). Thus the development of the real-time RTPCR or quantitative PCR (QPCR) for assessing accumulation of transcripts became another important tool for gene expression studies in the last decade and offered multiplexing capabilities of four to five genes simultaneously. Notwithstanding the sensitivity and relative accuracy of the QPCR, the limited number of genes which could be studied at a single time prevented global snap-shots of several genes simultaneously. To understand the up- and down-regulation of genes in a holistic perspective there was a requirement for looking at the expression of as many genes as possible; the so-called gene expression or transcriptome profiling approach. In addition to being generally sensitive, the developed methods for expression profiling were also high-throughput.

Currently there are three main large-scale high-throughput expression profiling techniques based on PCR, hybridization or DNA-sequencing. Hybridization strategy-reliant approaches include cDNA and oligonucleotide microarrays (Schena et al. 1995; Lockhart et al. 1996), while sequencing based approaches include the serial analysis of gene expression (SAGE) (Velculescu et al. 1995) and Massively Parallel Signature Sequences (MPSS) (Brenner et al. 2000). PCR-based approaches, which are also gel-based RNA fingerprinting methods, include Differential Display PCR

(DDRT-PCR) (Liang and Pardee 1992) and cDNA-Amplified Fragment Length Polymorphism (cDNA-AFLP) (Bachem et al. 1996). Over the years there have been several modifications or improvement of these methods and variants of the approaches are now used. Choice of the appropriate method is generally dictated by the resources available, the desired level of information anticipated based on the objectives of the project and the time-span to attain those objectives. It would therefore be useful to briefly review some of the benefits and challenges of some of these gene expression profiling approaches (Table 3.1).

3.6 Advantages and Drawbacks of Gene Expression Profiling Techniques

The complexity and resource-requirements of some of the available gene expression profiling approaches can be prohibitive for laboratories limited in terms of equipment and resources to conduct such profiling experiments (Table 3.1). It is therefore essential that resource allocation and amenities be judiciously managed. Approaches based on methods such as DDRT-PCR, cDNA-AFLP and variants thereof are gel-based transcript profiling systems dependent on electrophoretic fingerprinting of amplified cDNA fragments. There is no requirement for prior sequence knowledge for both DDRT-PCR and cDNA-AFLP techniques, and therefore these techniques allow for gene discovery. Only standard instrumentation requirements and low costs make these two approaches appealing for resource-limited laboratories. There are numerous published reports on the successful use of these methods (for reviews see Kuhn 2001; Green et al. 2001; Sarita and Yu 2006). Although a large number of mRNA species can be analyzed using DDRT-PCR, a high proportion of false positive putative differential display candidates are also generated. Furthermore, low abundance

Table 3.1 Comparison of commonly employed expression profiling approaches

Gene expression profiling techniques	PCR-based	Hybridization-based	Sequence-based
Examples	DDRT-PCR cDNA-AFLP	Macroarrays Microarrays	SAGE MPSS
Sensitivity	High	Moderate	High/moderate
Specificity	High	Low	High
Measure of expression level	Relative	Relative	Absolute
Prior sequence information	No	Yes	No
Gene discovery	Yes	No	Yes
Database storage and numerical manipulation	No	Yes	Yes
Molecular resource requirement	No	Yes	Yes
Labor intensity	High	Low	High
Cost	Low	High	High

transcripts are generally undetected. It is time-consuming since duplicate colony-lift screening is necessary. As opposed to the annealing temperatures used in DDRT-PCR, in cDNA-AFLP stringency and reproducibility can be improved by using relatively high annealing temperatures. Due to the sensitive and specific nature of the cDNA-AFLP technique, low-abundant transcripts can be detected as well as slight differences in transcriptional activity. The cDNA-AFLP method is based on selective amplification of cDNA fragments generated by restriction endonucleases, electrophoretic separation of the products and comparison of the band patterns between, for example, treated and control samples. Unequal distribution of restriction sites in cDNA fragments negatively affects the performance of cDNA-AFLP. Some transcripts are represented by more than one fragment, while others may not be detected (Breyne et al. 2003), leading to superfluous information and reduction in genome-wide coverage of the screen, respectively (Weiberg et al. 2008). Because of such limitations and technical requirements for cDNA-AFLP analysis, DDRT-PCR, has received more interest for expression profiling than cDNA-AFLP. Data obtained from both approaches are not quantitative in nature for database build-up and storage and for further numerical manipulations.

Hybridization-based approaches have the advantages of generating large amounts of data in a very short time, with significantly lower labor requirement. Quantitative data obtained from these techniques are suitable for database building and subsequently available for statistical manipulation of data. However, microarray analysis of gene expression is only feasible when extensive sequence information and/or cDNA libraries are available. Therefore, for plant genomes with limited sequence information, large-scale expression profiling using such methods may not be possible. Microarrays are sensitive to cross-hybridization noise limiting detection of low abundance transcripts. Compared to microarrays, cDNA-AFLP increases the resolution of expression pattern detection starting with smaller amounts of mRNA and is valuable for tissues from which the amount of RNA that can be isolated is sometimes technically challenging, as for example in anthers, pollens, ovules, fruits and seeds of several plants. Compared to microarrays, DNA sequencing-based methods such as SAGE and MPSS do not require prior sequence knowledge. Potentially novel mRNAs can also be identified using the sequence-based strategies. However, the SAGE approach is expensive, time- and labor-intensive, and prone to sequencing errors. MPSS, on the other hand, has the advantage over SAGE in not having to purify and sequence each clone since cDNAs are directly cloned onto beads and sequenced in parallel, saving significant time and effort, and thereby allowing greater in depth profiling of gene expression. Based on the currently available profiling technologies, MPSS offers the possibility of producing comprehensive information sets for use in hypothesis formulation and computational modeling to answer biological questions (Reinartz et al. 2002). An important advantage of the MPSS approach compared to SAGE is the possibility of generating more than 1,000,000 signatures rapidly (Reinartz et al. 2002). Furthermore, the MPSS tags are longer than most SAGE tags (reviewed in Coughlan et al. 2004). A major limitation of the technology is that it requires special equipment only available through Lynx Therapeutics, Inc., Hayward, California, USA.

3.7 Expression Profiling for Plant Improvement

Expression profiling data can be of significant value for plant breeding purposes in several ways. Coordinated expression of genes temporally, spatially, developmentally and in response to environmental stimuli affects the performance of a plant positively or negatively. Gene expression affects various physiological and biochemical processes culminating in the observable or discernible biological end results. Therefore, understanding of gene expression under various conditions is critical to understand biological processes for maximizing plant performance. Because of the large number of genes expressed in higher eukaryotic genomes, powerful tools have been developed to assess the overall pattern of gene expression, generally giving an indication of up- or down-regulated genes or absence of expression. Such studies will allow plant breeders to eventually select for combinations of expression profiles that lead to superior performance in their breeding lines. The advantage of currently available profiling methodologies, whether it is a "one gene at a time approach" or a multi-gene study, is that plant breeders can use the information for selection strategies under similar conditions to those used for generation of the gene expression profiles. For example, environmental effect on expression profile can be related to other traits that the breeder is interested in selecting.

Plant growth and productivity are affected by various biotic stresses such as herbivores, insects and pathogenic fungi, and abiotic stresses such as drought, salinity, cold, heat, anaerobiosis and heavy metals. Unlike animals, the sessile nature of higher plants prevents flight from the source of attack. However, plants have generally developed adaptive mechanisms, from perception of attack to defense response initiation by induction of a cascade of events composed of molecular signals and interacting networks, to cope with the perceived perturbation in their normal functioning. Gene expression profiling has identified genes expressed during plant infection or disease establishment, thereby elucidating the molecular basis of these processes and identifying genes that could help to inhibit or prevent further pathogen attack. During plant-pathogen interaction studies, gene expression profiling techniques have allowed the identification of virulence factors, genes involved in cell death, signaling and resistance for potential application in plant breeding (Table 3.2). Using DDRT-PCR, differentially expressed transcripts were indentified from the fungal pathogen *Botrytis cinerea* and its tomato host plant (Benito et al. 1996). Similarly, a cDNA-AFLP approach for studying differential transcript induction as a result of sugar cane-smut fungus interaction identified genes involved in oxidative burst, defense response, and auxin and ethylene pathways, suggesting that genes for oxidative burst and lignin pathways played major roles in the first 72 h of infection (Lao et al. 2008). In a study on resistance to the fusarium mycotoxin, deoxynivalenol (DON), microarray and real-time PCR analyses of progeny from a cross between a resistant and a susceptible genotype showed that ten toxin-responsive transcripts were associated with inheritance of DON resistance as well as to the resistance QTL, Fhb1 (Walter et al. 2008). Such information would be of value for breeders to select for accumulation patterns of these particular transcripts in addition to selecting for the resistance QTL and the DON-resistance phenotype.

Table 3.2 Expression profiling to identify differentially expressed genes in plants under biotic stresses

Techniques used	Plant	Biotic stressors	References
DDRT-PCR	*Rostrate sesbania*	*Azorhizobium caulinodans*	Goormachtig et al. (1995)
	Tomato	*Botrytis cinerea*	Benito et al. (1996)
	Chickpea	*Ascochyta rabiei*	Rajesh et al. (2003)
	Cotton	*Verticillium dahliae*	Zwiegelaar and Dubery (2006)
	Wheat	*Puccinia striiformis*	Bozkurt et al. (2007)
cDNA-AFLP	Sugarcane	*Sporisorium scitamineum*	Lao et al. (2008)
	Wheat	*Fusarium graminearum*	Steiner et al. (2008)
	Rice	Rice yellow mottle virus	Ventelon-Debout et al. (2008)
	Tomato	*Cladosporium fulvum*	Zhu et al. (2008)
	Medicago truncatula	Caterpillar	Darwish et al. (2008)
Super SAGE	Pepper	Tobamovirus	Hamada et al. (2008)
Long SAGE	Tomato	Cucumber mosaic virus/ D satellite RNA	Irian et al. (2007)
Microarray	Arabidopsis	*Alternaria brassicicola*	Schenk et al. (2000)
	Wheat	*Fusarium* mycotoxin	Walter et al. (2008)

As with gene expression profiling studies for biotic stresses, the influence of abiotic stresses on expression patterns of genes in plants have also been extensively studied (Table 3.3). Some of the abiotic stress factors such as heavy metal toxicity or salinity to some extent are more defined, while others such as cold, heat and drought are less defined and more variable even during the course of a day, and can lead to interactions which generate more complex gene expression profiles in plants exposed to these stresses. Mimicking such environmental states for gene expression profiling studies has therefore been given significant importance to reflect as close a situation as possible to actual environments. Thus there have been artificial systems established for triggering automated changes in factors such as temperature, drought, light intensity and quality and humidity.

Due to losses encountered as a result of the various abiotic stresses, there have been extensive efforts by breeders to develop tolerant varieties. However, the complex nature of the inheritance of abiotic stress tolerance has hindered absolute success. Thus numerous expression profiling studies have addressed this issue to facilitate selection for abiotic stress tolerant plants. While the practical implications of these studies have not yet been fully realized, some examples are provided where immediate implications or potential applications are already discernible for crop improvement. Using DDRT-PCR in barley, several partial cDNA clones corresponding to cold-, drought- and ABA-induction were identified and one of the cDNAs was mapped to a region on chromosome 7H previously associated with osmotic adaptation (Malatrasi et al. 2002). In another recent study with microarrays in rice differentially expressed genes from drought tolerant and sensitive cultivars

Table 3.3 Expression profiling to identify differentially expressed genes in plants under abiotic stresses

Abiotic Stresses	Plants	Techniques used	References
Drought	Barley	DDRT-PCR	Malatrasi et al. (2002)
	Cotton	DDRT-PCR	Maqbool et al. (2008)
	Festuca mairei	cDNA-AFLP	Wang and Bughrara (2007)
	Maize	Oligonucleotide Microarray	Zhuang et al. (2008)
	Rice	Oligonucleotide Microarray	Degenkolbe et al. (2009)
Salinity	Rice	DDRT-PCR	Zhang and Chen (1996)
	Foxtail millet	cDNA-AFLP	Jayaraman et al. (2008)
	Arabidopsis	Microarray	Krishnaswamy et al. (2008)
	Rice	Microarray	Huang et al. (2008)
Chilling	*Citrus unshiu*	DDRT-PCR	Lang et al. (2005)
	Maize	DDRT-PCR	Liu et al. (2007)
	Alfalfa	cDNA-AFLP	Ivashuta et al. (1999)
	Cucumber	cDNA-AFLP	Lu et al. (2005)
	Sunflower	cDNA Microarray	Fernandez et al. (2008)
	Arabidopsis	SAGE	Jung et al. (2003)
	Arabidopsis	SAGE	Robinson and Parkin (2008)
Heat	Arabidopsis	DDRT-PCR	Visioli et al. (1997)
	Senecio × hybridus	DDRT-PCR	Yu et al. (2007)
	Wheat	cDNA-AFLP	Li et al. (2007)
	Cowpea	cDNA-AFLP	Simoes-Araujo et al. (2008)
	Rice	Microarray	Yamakawa et al. (2007)
Metal Toxicity			
Aluminium	Rice	DDRT-PCR	Zhang et al. (2007)
	Rice	cDNA-AFLP	Mao et al. (2004)
	Wheat	cDNA-AFLP	Houde and Diallo (2008)
	Medicago truncatula	Oligonucleotide Microarray	Chandran et al. (2008)
Cadmium	Mustard	cDNA-AFLP	Fusco et al. (2005)
Boron	*Gypsophila perfoliata*	Microarray	Unver et al. (2008)
Ozone	*Betula pendula* Roth	cDNA Microarray	Kiiskinen et al. (1997)
UV-B irradiation	European Beech	DDRT-PCR	Ernst et al. (2001)
Anaerobiosis	Rice	DDRT-PCR	Chen et al. (2007)

were mapped to a drought tolerant QTL identified from published QTL data based on G × E interactions for drought tolerance (Degenkolbe et al. 2009).

Besides the applications of expression profiling approaches for breeding biotic and abiotic stress tolerance, quality and developmental attributes have also been targeted. Again a vast repertoire of data is available and incorporation of such information into breeding strategies is just beginning to emerge. For example, microarray studies based on potato tuber-specific sequences have revealed strong up- or down-regulation of several transcripts during development, and led to identification of

differential gene expression related to tuber development, starch biosynthesis and tissue specificity (Kloosterman et al. 2005). Identification of important regulatory sequences from such studies is likely to allow more targeted breeding for development of specialty starches by modifying metabolic routes (Kloosterman et al. 2005). Similarly, using microarray studies differences in gene expression data were related to malting quality traits in barley and suggested to be of value for future barley malting quality improvement (Potokina et al. 2004).

3.8 eQTL, Gene Networks and Pathways

Plant breeders have always relied on genetic variability for selection in their breeding programs. As mentioned earlier, the difficulty over time for identification of such variability increased as a result of dwindling germplasm resources and also limited molecular level understanding of plant functions. The paradigm shift from qualitative to quantitative traits in molecular mapping and marker development heralded a major initiative towards global perspective on the influence of quantitative traits and the quantitative trait loci (QTL) association with complex traits and the possibility of identifying and cloning candidate genes by map-based cloning. Microarray studies recognize up- and down-regulated genes among individuals under different conditions and/or plant developmental stages. It became evident that this expression profiling approach was a powerful tool to comprehensively understand the global patterns of gene expression. To further improve on this perception, a new concept of "genetical genomics" was put forward to relate genetic variation due to QTL in a segregating population to genome-wide expression profiling data, and therefore identify the proportion of the variation in gene expression actually associated with such QTL (Jansen and Nap 2001). Validation for the genetical genomics concept has been shown in mouse, human and maize, wherein expression values from transcriptome data analyses of these organisms were treated as quantitative traits and used for linkage analyses and standard interval mapping along with markers specific for these organisms (Schadt et al. 2003). The study also demonstrated that generally the identified gene expression QTL (eQTL) was able to account for the transcription variation of the genes associated with a trait.

Besides the advantage of eQTL in relating expression data as a quantitative trait for association with phenotypes, it was also suggested that the proportion of expression variation that actually maps to the genes themselves would qualify as *cis*-acting factors and those mapping to other regions as *trans*-acting factors (Jansen and Nap 2001). The possibility of mapping *cis*-eQTL and *trans*-eQTL was demonstrated in a mouse system, wherein generally eQTL with high LOD scores corresponded to the former and those that are moderately significant corresponded to the latter (Schadt et al. 2003). Such *cis*- and *trans*-eQTLs have also been identified in a recombinant inbred line population of *Arabidopsis* using microarray analysis and eQTL mapping of transcript abundance as quantitative expression traits (West et al. 2007). The study, similar to the findings of Schadt et al. (2003), identified eQTL hotspots.

These eQTL hotspots are the result of a single upstream polymorphism affecting expression of a large number of distant genes, which are mapped at the same locus (Breitling et al. 2008). Such eQTL hotpots have also been reported in maize in relation to the cell wall digestibility trait, with the eQTL being co-localized to the QTL cluster (Shi et al. 2007). Thus, besides being able to map transcript abundance as a function of quantitative expression trait and possibly more precisely identifying candidate genes, the eQTL approach will enable plant breeders to further refine selection for traits of interest based on significant eQTL, indirectly selecting for combinations of genes expressing at levels allowing for maximal genetic potential.

Genetic interactions have been common themes since the beginning of the unraveling of genetic laws. With advances in quantitative and computational genetics and QTL mapping, dissection of more complex interactions has been possible. Networks and pathways have been recognized mostly based on biochemical interactions and their effects on biological functions. However, with data from genome-wide expression as well as eQTL mapping, there is a new era in functional genomics with attempts to elucidate gene networks and pathways. Genetic network has been defined as being constituted of a group of genes, wherein individual genes can affect the expression of other genes (Wagner 2002). It has become evident that such networks and pathways are operating at the expression level and generally lead to variation in phenotypes, which have been unexplainable. Due to the vast array of data from expression profiling studies, a systems approach is now being taken to understand the interactions and pathways. Systems biology attempts to study multi-component effects as applied to functional biology and the systems approach has gained further impetus with the possibility of disrupting specific functions (Ideker et al. 2001). To further advance such an approach to function analysis, modeling and computational algorithms are being applied, and it is expected that a better understanding of gene regulatory networks will be obtained (reviewed in Long et al. 2008).

3.9 Transgene Technology and Gene Expression Analysis

The basis for several developments in gene transfer to plants rested solely on the availability of a tissue culture and plant regeneration system. It is still a requirement for the majority of plant species to be genetically altered by exogenous gene transfer. Tissue culture techniques became fairly well established over the major part of the twentieth century. In the last three decades of the twentieth century, extensive efforts were made to optimize and develop highly efficient and reproducible tissue culture systems, especially for economically important crop plants to make them amenable to genetic transformation. Generally, dicotyledonous plants responded better to in vitro culture than monocotyledonous plants. Thus, plant tissue culture approaches have over the latter quarter of the last century focused on sources of explant tissues for culture, culture media composition and plant growth regulators, methods of shoot regeneration and other miscellaneous factors (Ganeshan et al. 2002). Efforts are still on-going for several crop plants for maximization of regeneration efficiency to

concurrently increase gene transfer efficiency. There is an exhaustive list of explant sources from various plant species that have been tested for their ease of isolation and response to culture for efficient regeneration (Table 3.4). The requirements for the successful transformation of plants therefore relied extensively on plant tissue culture systems which could be fairly easily established, showed high regeneration efficiency and finally also exhibited high transformation efficiency. The latter, however, although more directly related to the DNA delivery method, is indirectly affected by the choice of the recipient tissue/cells for the gene transfer process, per se.

To understand the role of plant tissue culture in genetic transformation, a brief overview highlighting history and principles governing the process would also be useful in the context of functional genomics in crop improvement. Among all living organisms, the ability to re-program plant cells/tissues to regenerate whole new plants is unique. The concept of totipotency attributed to Gotlieb Haberlandt and his work and postulates in 1902 (for details and translation of Haberlandt's original work see Krikorian and Berquam 1969) laid the foundation for the evolution of plant tissue culture as an area of research. It is now well-established that a vast number of explant systems can be in vitro cultured for the regeneration of plants. However, the pathways which lead to the generation of entire new plants from the various explants and the manner in which such pathways can be influenced exogenously need to be understood (Fig. 3.2).

With regards to transformation approaches, tissue culture-induced variation is not desirable. As mentioned earlier, somaclonal variation was identified as an added source of genetic variability (Larkin and Scowcroft 1981), but such variation would not be acceptable in transgenic plants. Furthermore, variation induced as a result of in vitro culture can also impact the regenerated plant negatively (for reviews see Karp 1995; Vazquez 2001; Filipecki and Malepszy 2006), which would preclude the added advantage of transgenic events with respect to genetically transformed crops. Generally, it is during the callus phase that variation has been suggested to occur due to the uncontrolled cell division (Phillips et al. 1994). It would therefore be advantageous to circumvent this stage if direct shoot induction from the explant were permissible. However, it should be mentioned that variation is also induced due to the effects of growth regulators in culture, irrespective of a callus phase (Karp 1995) or

Table 3.4 Examples of different types of explants explored for their tissue culture responses in monocotyledonous and dicotyledonous plant species for subsequent use in transformation

Monocotyledons	Dicotyledons
Protoplasts	Protoplasts
Anthers/Microspores	Anthers/microspores
Immature embryos	Cotyledonary segments
Immature scutella	Cotyledonary nodes
Leaf-bases	Leaf tissue segments
Cell suspensions	Hypocotyls
Immature inflorescences	Shoot apices
Shoot apical meristems	Embryo segments
Inflorescences	
Mature embryos	

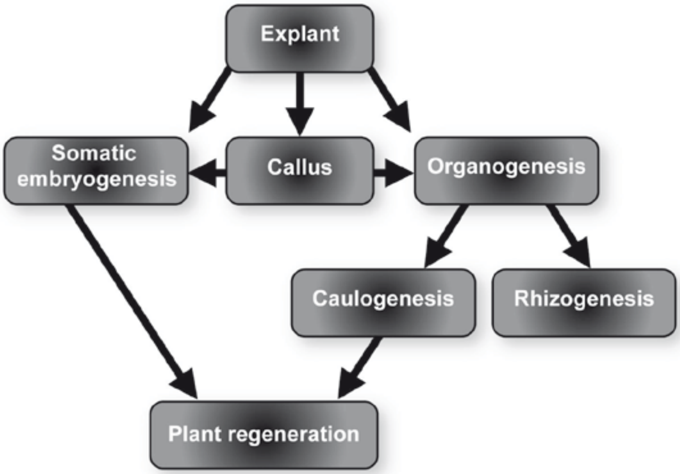


Fig. 3.2 Overview of a generalized tissue culture approach showing de-differentiation routes from the explant

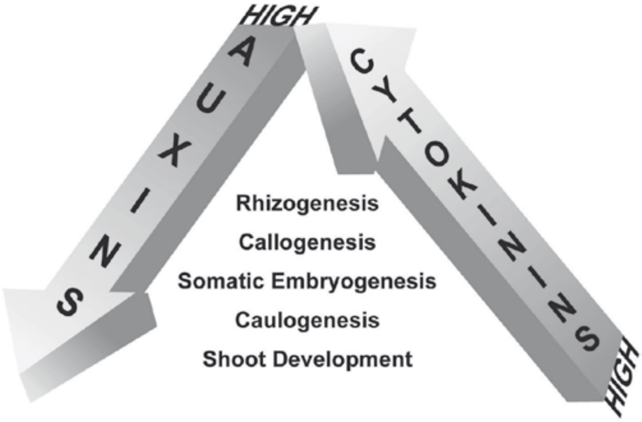


Fig. 3.3 Genera influence of auxins to cytokinins ratio in cell fate determination. Increasing or decreasing the ratio of auxins to cytokinins during the different phases of the tissue culture process can lead to a variety of ontogenic developments from the explant tissue

even in clonally propagated plants, and it is generally recognized that besides culture environment and explant source, prolonged duration of culture, genotype, explant age and ploidy level can contribute to variation (reviewed in Rani and Raina 2000). Therefore, a certain amount of preliminary tissue culture studies need to be conducted before initiating genetic transformation projects, especially if there is a lack of prior published information in the literature. In a general scheme of tissue culture responses from explants it is well established that the ratio of auxins to cytokinins plays an important role in the ontogenic developmental phases (Fig. 3.3).

Thus high levels of auxins can lead to induction of roots (rhizogenesis) or callus induction (callogenesis), while high levels of cytokinins can lead to shoot initiation (caulogenesis) and further shoot development. In many plant species, plant growth regulators such as thidiazuron (TDZ), with both auxin- and cytokinin-like properties, have been shown to induce axillary and/or adventitious shoots or somatic embryos (Huetteman and Preece 1993; Lu 1993; Murthy et al. 1998). More recently TDZ has also been shown to induce direct multiple shoots from cereal mature embryos, without an intervening callus phase and with no adverse effects in agronomic performance of regenerated plants (Ganeshan et al. 2003; Ganeshan et al. 2006).

3.10 Choice of Target Tissue for Gene Delivery

The choice of the explant for genetic transformation depends largely on the tissue culture responsiveness of the explant and also on the gene delivery method. For example, the harsh physical nature of gene delivery methods such as the particle gun is likely to cause extensive damage to soft and delicate tissues, thereby affecting subsequent regeneration and recovery of transgenic plants. With *Agrobacterium* mediated transformation the elimination of the bacteria after co-cultivation and infection is achieved by use of antibiotics and such antibiotics can interfere with regeneration from the infected explants. Furthermore, an important consideration in genetic transformation is the use of a plant tissue culture system which minimizes or prevents occurrence of chimerism (Gaj 2004). Plant chimerism is essentially the occurrence of cells of different genetic make-up within the same plant. Thus transformation of a single cell which would lead to production of an entire plant either through organogenesis or embryogenesis would be preferable. However, in the tissue culture and transformation methods currently available chimerism may not always be preventable. From a shoot apical meristem (SAM) perspective, plant chimeras have been described as the existence of cells of different genetic make-up within the meristems which eventually contribute to the formation of the organs and tissues of the plant (Burge et al. 2002). In order to understand chimerism with regards to transformation, an example using the SAM is described below, and it would therefore be valuable to briefly review the organization of the SAM.

The shoot apical meristem is generally distinguished into three regions: the central zone (CZ), the rib zone (RZ) and the peripheral zone (PZ) (Fig. 3.4) (for reviews see Szymkowiak and Sussex 1996; Barton 1998; Bowman and Eshed 2000; Burge et al. 2002; Tooke and Battey 2003; Kwiatkowska 2008). The outermost layer of the apical meristem dome consists of the tunica, made up of the L1 and L2 layers, referred to as the epidermal and sub-epidermal layers, respectively (Fig. 3.4). Cell division in the tunica is anticlinal. In the L3 layer, referred to as the corpus, cell division occurs in different planes. The contribution of the different apical meristem zones is such that the peripheral zone contributes to lateral organs such as the leaves, the rib zone contributes to most of the stem and the central zone, referred to as a

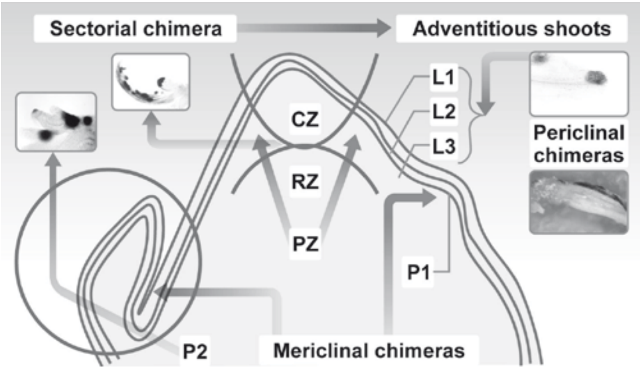


Fig. 3.4 Cell zones in the shoot apical meristems and types of chimerism which can result from transformation of shoot apical meristems region. L1 and L2 layers – tunica; L3 layer – corpus; CZ – central zone; RZ – rib zone; PZ – peripheral zone; P1 and P2 – leaf primordia

reservoir of stem cells, replenishes the PZ and RZ while maintaining the CZ itself (Bowman and Eshed 2000).

The three types of chimeras generally recognized based on the occurrence of the genetically different cell types within different layers are the periclinal, mericlinal and sectorial chimeras (Marcotrigiano 1997). Periclinal chimeras are recognized when one or more entire cell layers are genetically different from another cell layer, as for example, when the L1 layer of the apical meristem is genetically different. Mericlinal chimeras occur when the genetically different cells are found in part of a cell layer. Sectorial chimeras are those with sectors of all cell layers being genetically different. Such chimeric types can also be distinguished in transformed tissues. Using particle bombardment and a construct carrying the *Gus* gene, transformation of wheat mature embryos can be used to exemplify the occurrence of transgenic chimeras (Fig. 3.4). This system of mature cereal embryo culture on a thidiazuron containing medium has been shown to produce multiple shoots without an intervening callus phase (Ganeshan et al. 2003; Ganeshan et al. 2006). Upon performing GUS histological assays, it has been possible to distinguish a periclinal chimeric region wherein only the epidermal surface layers exhibited blue coloration (Fig. 3.4). Similarly mericlinal chimeras can arise if the integration of the *Gus* gene occurred at the leaf primordial initiation region. Sectorial chimeras can be recognized if the CZ or PZ become transformed. With the production of adventitious shoots from such transformed sectorial chimeras an entire transformed shoot can arise. Although the types of chimeric transformants described here have been exemplified using the SAM, such chimeras can also occur from other types of explants used for genetic transformation. To increase transformation events, it is essential to recognize occurrence of chimerism and promote sectorial chimeras. This can generally be achieved by judicious antibiotic or other selection strategies.

3.11 Gene Transfer Methods

Gene transfer to plants has been a challenging area of research. The two main methods of gene transfer to plants are a biological method involving *Agrobacterium*, or a physical method involving the particle bombardment. There are several other methods of gene transfer to plants such as electroporation, micro-injection, liposome-mediated, silicon carbide whiskers, bio-active beads and more recently there has been an interest in nanoparticles. *Agrobacterium*-mediated gene transfer to generate a transgenic plant was first reported in 1983 (Barton and Chilton 1983; Fraley and Horsch 1983). However, it was initially observed that dicotyledonous plants were more amenable to transformation by *Agrobacterium* as compared to monocotyledonous plants. In the early 1990s there were several attempts to transform some of the economically important monocotyledonous plants such as cereals using *Agrobacterium*, with rice being a prime candidate. The first successful report on the stable *Agrobacterium*-mediated transformation of Japonica rice was in 1994 (Hiei et al. 1994). Subsequently there have been several reports on the transformation of indica as well and javanica rice (reviewed in Shrawat and Lorz 2006). It is now possible to routinely transform rice using *Agrobacterium*. Although successes with other cereals such as wheat, maize and barley have been possible in recent years, the efficiency is still not adequate for the demands of high-throughput functional genomics projects. For plant species not amenable to transformation by *Agrobacterium*, alternative gene transfer methods have been sought. However, the major impediment has always been to find another biological entity capable of transferring DNA into the plant cell or finding a physical method to insert DNA.

Plant viruses have for long been considered a biological alternative. It has been known since the 1920s that certain virus-infected crop plants became resistant to the same plant viruses or closely related strains of those viruses (Mckinney 1927, 1929, 1937). The phenomenon of cross-protection used to describe such acquired resistance (Fulton 1986) resulted in the concept of pathogen-derived resistance (PDR) for engineering such resistance into crop plants (Sanford and Johnston 1985; Grumet et al. 1987). The validity of the concept was demonstrated with the production of transgenic tobacco plants carrying the tobacco mosaic virus coat protein gene and exhibiting delayed symptoms upon infection with the virus (Abel et al. 1986). The operation of a post-transcriptional gene silencing (PTGS) mechanism was subsequently attributed to the delay or absence of symptoms (Ratcliff et al. 1997). Although the transfer of the gene was achieved using *Agrobacterium*, the silencing aspect of the approach started gaining attention and led to virus-induced gene silencing (VIGS) to describe transgenic plants which recovered from virus infection (van Kammen 1997). VIGS has become a powerful tool to down-regulate genes of interest in high-throughput functional genomics studies (Baulcombe 1999). The advantage of the VIGS approach is that gene transfer is not necessary since it has been shown that rubbing of in vitro transcribed viral RNA containing a phytoene desaturase (PDS) gene on leaves of *Nicotiana benthamiana* caused down-regulation of the PDS (Kumagai et al. 1995). Therefore a transient virus-mediated system is

very valuable for rapidly studying the silencing or down-regulation of a number of genes. The VIGS mechanism has also been shown to work in monocotyledonous plant using the barley stripe mosaic virus system (Holzberg et al. 2002).

During the early research in *Agrobacterium* crown gall tumor formation, studies showed that crown gall induction ability could be transferred from virulent to avirulent strains of *A. tumefaciens* and also to *A. radiobacter*, *A. rubi* and *Rhizobium leguminosarum* (Klein and Klein 1953). Subsequently it was demonstrated that crown gall formation occurred by transfer of the Ti plasmid to avirulent strains of *Agrobacterium* and to *Rhizobium* (Hooykaas et al. 1977) as well as *Phyllobacterium myrsinacearum* (Veen et al. 1988). More recently Broothaerts et al. (2005) have reported on the transfer of disarmed Ti plasmid to diverse species of bacteria such as *Rhizobium*, *Sinorhizobium meliloti* and *Mesorhizobium loti*, which were used for the transformation of tobacco, rice and *Arabidopsis*. It may therefore be possible for such alternative bacteria to broaden the host range for T-DNA transfer and also for transformation of plant species that have been so far relatively recalcitrant to *Agrobacterium*. This strategy using other bacterial species is available as an open-source-modeled license to all researchers. Details are available at the Bioforge project (http://www.bioforge.net) and at Biological Innovation for Open Society (BIOS; http://www.bios.net).

The most common physical method used for DNA transfer to plants is the microprojectile bombardment method, also known as the biolistic method (biological ballistic) invented by Sanford et al. in 1987. Microprojectile bombardment method circumvents the host range problems associated with *Agrobacterium* and the physical barrier of the plant cell wall was not an issue for DNA entry (Weissinger et al. 1987; Klein et al. 1988a, b). The underlying operation of the particle bombardment approach involves the acceleration of micro-particles carrying DNA to high velocity such that they are able to penetrate the cell wall and deliver the nucleic acid into the nucleus of the cell (Sanford 1990). There have been several modifications to the particle bombardment apparatus over the years including use of nitrogen gas (Morikawa et al. 1989), compressed air (Iida et al. 1990), air gun (Oard et al. 1990) or helium gas (Finer et al. 1992) to deliver the micro-particles. Some of the general considerations in the particle bombardment process for efficient transformation are the types of tissues to be used, the flight distance and velocity of the micro-particles. Due to damage to tissues with this method, regeneration subsequent to transformation can be affected. Therefore during optimization efforts, all the parameters are tested such that there is minimal damage to the tissues, while obtaining a high transformation efficiency. Due to the ease of using the biolistic method of transformation there are numerous reports of transformed plant species including many agricultural crops. The challenge for functional genomics studies, however, has been the efficiency of transformation with the microprojectile bombardment method being generally low compared to *Agrobacterium*-mediated transformation of plants such as *Arabidopsis* and rice. Furthermore, because of the harsh nature of delivery of the DNA, there are multiple insertion events which can complicate assessment of the effect of the transgene of interest. Besides the somaclonal variation effects mentioned earlier, the transformation process itself can cause poor agronomic performance of

the transgenic plants. Also, position effect of the transgene can affect its expression and stability (for review see Filipecki and Malepszy 2006).

Other physical methods of gene transfer include electroporation, liposomes, silicon carbide whiskers and recently nanoparticles. Electroporation of plants has employed short electric pulses to deliver DNA into plant protoplasts, cells or tissues. Protoplasts, due to the lack of the cell wall, were more amenable to uptake DNA. Unfortunately, regeneration of fertile transgenic plants from electroporated protoplasts is generally very difficult and this approach has not been widely used. Attempts at electroporating plant cells and tissues such as pollens, embryos and callus have been made with varying success. Although feasible, the electroporation-mediated transformation requires extensive optimization efforts with low transformation efficiencies. Similarly, the use of liposomes for transformation would require the protoplast route for delivery of DNA. Among the physical delivery methods, the silicon carbide whisker-mediated DNA delivery appears to be an appealing alternative due to its inexpensiveness and simplicity. However, as with the biolistic method, complex transgene integration patterns have been reported (Petolino et al. 2000). Thus for functional genomics studies these physical methods need major improvements, especially in terms of high efficiency transformation, to be of consequence. The recent trends in nanoparticle research may open up new possibilities in high efficiency plant transformation. Many of the breakthroughs for plant transformation using nanoparticles may come from studies conducted in animal and mammalian cell systems. There have been extensive efforts to use nanoparticles in these systems for molecule and drug delivery (Brigger et al. 2002; Yih and Al-Fandi 2006). Recently, using mesoporous nanoparticles (MSNs) containing a trans-activator for a GFP expression construct, it was shown that expression of the GFP gene could be induced *in vivo* (Torney et al. 2007). The trans-activator was filled within the MSNs and the GFP construct coated onto gold particles was used to cap the MSNs, before delivery into plant cells by microprojectile bombardment. Further research into such targeted and controlled delivery of nanoparticles may be of value for enhancing transformation efficiency in plants for functional genomics studies.

3.12 Transgene Approach to Functional Genomics and Plant Breeding

The application of transgene technology to study gene function has become essential in the post-genomic era and more importantly to aid plant breeders with selecting for performance enhanced combinations of genes in newer varieties. Therefore, the transgenic approach has been used in a variety of ways to assess gene function in plants (Fig. 3.5). From gene inactivation to over-expression studies to insertional mutagenesis and traps and tags, gene transfer is an underlying requirement. Even though initially use of gene transfer was primarily aimed at production and commercial release of transgenic plants, its use for gene function assessment was nonetheless realized. For several crops, especially cereals, it was found that the

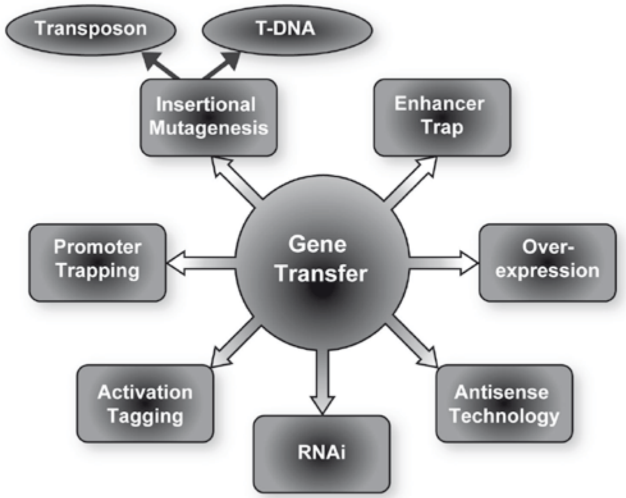


Fig. 3.5 Gene transfer and different approaches for functional characterization of genes

transformation efficiency was high in certain model cultivars, which were not necessarily agronomically of outstanding performance (e.g., Bregitzer et al. 2008). Therefore there was a need for plant breeders to be able to introgress such transgenic events into commercially viable cultivars or even to eliminate undesirable effects in a transgenic plant by backcrossing. For example, transgenic barley carrying a transgene coding for an engineered heat-stable (1,3–1,4)-β-glucanase showed reduced yield among the transgenic lines, but showed yield improvements after the transgene was introgressed into another genotype (Horvath et al. 2001). While this approach is useful for transgenes of commercial interest to be in the genetic background of elite breeding lines or varieties, for functional analyses of genes this would be time-consuming and resource-demanding.

3.13 Prospects and Challenges

The advances in functional genomics tools over the last decade have widened the scope for newer technologies to provide additional resources for accelerating plant breeding programs. Advances in gene expression profiling approaches have led to more efficient and high-throughput data being generated. An exhaustive amount of data is being generated from expression profiling techniques like microarrays, and bio-informatics analyses combine such expression data with QTL data to generate eQTL profiles to give a global picture of the genetic basis of important agricultural traits. It is highly likely that over the next few years breeding programs will begin to actually incorporate such knowledge for the development of new varieties.

Furthermore it is foreseen that with the combination of eQTL mapping and identification of more refined gene networks, more precise candidate genes responsible for traits of interest are likely to be cloned.

The challenge for functional genomics studies in terms of gene expression profiling is likely to be in the form of experimental set-up and design. Generally most of the studies are conducted under controlled conditions, as for example when studying abiotic stresses. Appropriate stress factors are applied in quantifiable amounts and changes in gene expression, physiological and biochemical profiles are examined. However, for plants growing under natural conditions of environmental stresses extrapolation of results obtained using such studies can often be inconclusive, and therefore the expression profile data may be of limited value for breeding programs. For practical applications, considerable emphasis is placed on genotype × environment interactions for selection of stable performance across environments or within a particular environment. This also applies to biotic stresses where the nature of the resistance mechanism is quantitative. Recently a field trial with winter wheat was conducted to assess expression profiles of cold-regulated genes over a three year period (Ganeshan et al. 2009). In general, there was some agreement with growth chamber studies for gene expression, although variability over the three years was observed and attributed to soil temperature fluctuations. In another growth chamber study the choice of tissue for gene expression profiling was found to significantly influence expression of selected cold-regulated genes (Ganeshan et al. 2008), and therefore influenced biological interpretations and/or conclusions. These studies iterate the importance of proper experimental design and controls and also choice of appropriate materials for gene expression profiling studies.

A further challenge for functional genomics studies lies in the fact that gene transfer approaches for several economically important plants are still inefficient. The repertoire of genome sequence and expression data is accumulating at a feverish pace, but assessment of the functions of the genes in the relevant plants in a high-throughput manner has become a major challenge. It can be argued that model systems such as *Arabidopsis* and rice can be used, but the results obtained cannot always be extrapolated to the respective plants. Therefore major efforts are required to re-invest resources into developing more efficient and high-throughput transformation systems to assess gene function.

Notwithstanding these challenges, the decade ahead is likely to see more extensive contributions from systems biology, in silico and virtual expression profiling which are likely to propel functional genomics studies for the benefit of crop improvement. The combination of breeding schemes such as the diallel cross with genome-wide expression analysis for *cis*- and *trans*-regulatory variants (Kiekens et al. 2006) and incorporation of expression data for specific and general combining ability determination (Vuylsteke and van Eeuwijk 2008) are recent examples of direct application of classical plant breeding approaches integrating functional genomics data.

Acknowledgements Genome Canada, Genome Prairie, Canada Research Chairs, Canada Foundation for Innovation, and Natural Sciences and Engineering Research Council, are gratefully acknowledged for financial support of research in our laboratory.

References

Abel PP, Nelson RS, De B, Hoffmann N, Rogers SG, Fraley RT, Beachy RN (1986) Delay of disease development in transgenic plants that express the tobacco mosaic virus coat protein gene. Science 232:738–743

Ahloowalia BS, Maluszynski M, Nichterlein K (2004) Global impact of mutation-derived varieties. Euphytica 135:187–204

Bachem CWB, van der Hoeven RS, De Bruijn SM, Vreugdenhil D, Zabeau M, Visser RGF (1996) Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. Plant J 9:745–753

Baillie AMR, Rossnagel BG, Kartha KK (1993) In vitro selection for improved chlorsulfuron tolerance in barley (*Hordeum vulgare* L.). Euphytica 67:151–154

Barton KA, Chilton MD (1983) *Agrobacterium* Ti plasmids as vectors for plant genetic engineering. Methods Enzymol 101:527–539

Barton MK (1998) Cell type specification and self renewal in the vegetative shoot apical meristem. Curr Opin Plant Biol 1:37–42

Baulcombe DC (1999) Fast forward genetics based on virus-induced gene silencing. Curr Opin Plant Biol 2:109–113

Benito EP, Prins T, vanKan JAL (1996) Application of differential display RT-PCR to the analysis of gene expression in a plant–fungus interaction. Plant Mol Biol 32:947–957

Bidani A, Nouri-Ellouz O, Lakhoua L, Sihachakr D, Cheniclet C, Mahjoub A, Drira N, Gargouri-Bouzid R (2007) Interspecific potato somatic hybrids between *Solanum berthaultii* and *Solanum tuberosum* L. showed recombinant plastome and improved tolerance to salinity. Plant Cell Tissue Organ Cult 91:179–189

Blakeslee AF, Belling J, Farnham ME, Bergner AD (1922) A haploid mutant in the jimson weed *Datura stramonium*. Science 55:646–647

Borlaug NE (2002) Feeding a world of 10 billion people: The miracle ahead. In Vitro Cell Dev Biol Plant 38:221–228

Bowman JL, Eshed Y (2000) Formation and maintenance of the shoot apical meristem. Trends Plant Sci 5:110–115

Bozkurt O, Unver T, Akkaya M (2007) Genes associated with resistance to wheat yellow rust disease identified by differential display analysis. Physiol Mol Plant Pathol 71:251–259

Bregitzer P, Dahleen LS, Neate S, Schwarz P, Mancharan M (2008) A single backcross effectively eliminates agronomic and quality alterations caused by somaclonal variation in transgenic barley. Crop Sci 48:471–479

Breitling R, Li Y, Tesson BM, Fu JY, Wu CL, Wiltshire T, Gerrits A, Bystrykh LV, de Haan G, Su AI, Jansen RC (2008) Genetical genomics: spotlight on QTL hotspots. PLoS Genet 4(10): e1000232. doi: 10.1371/journal.pgen.1000232

Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo SJ, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J, Corcoran K (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol 18:630–634

Breyne P, Dreesen R, Cannoot B, Rombaut D, Vandepoele K, Rombauts S, Vanderhaeghen R, Inze D, Zabeau M (2003) Quantitative cDNA-AFLP analysis for genome-wide expression studies. Mol Genet Genomics 269:173–179

Brigger I, Dubernet C, Couvreur P (2002) Nanoparticles in cancer therapy and diagnosis. Adv Drug Del Rev 54:631–651

Broothaerts W, Mitchell HJ, Weir B, Kaines S, Smith LMA, Yang W, Mayer JE, Roa-Rodriguez C, Jefferson RA (2005) Gene transfer to plants by diverse species of bacteria. Nature 433:629–633

Buiatti M, Ingram DS (1991) Phytotoxins as tools in breeding and selection of disease-resistant plants. Experientia 47:811–819

Burge GK, Morgan ER, Seelye JF (2002) Opportunities for synthetic plant chimeral breeding: past and future. Plant Cell Tissue Organ Cult 70:13–21

Chandran D, Sharopova N, Ivashuta S, Gantt JS, VandenBosch KA, Samac DA (2008) Transcriptome profiling identified novel genes associated with aluminum toxicity, resistance and tolerance in *Medicago truncatula*. Planta 228:151–166
Chase SS (1949) Monoploid frequencies in a commercial double cross hybrid maize, and in its component single cross hybrids and inbred lines. Genetics 34:328–332
Chase SS (1952) Production of homozygous diploids of maize from monoploids. Agron J 44:263–267
Chen Q, Lynch D, Platt HW, Li HY, Shi Y, Li HJ, Beasley D, Rakosy-Tican L, Theme R (2004) Interspecific crossability and cytogenetic analysis of sexual progenies of Mexican wild diploid 1EBN species *Solanum pinnatisectum* and *S. cardiophyllum*. Am J Potato Res 81:159–169
Chen Y, Zhao S, Yan Q, Li Y, Wu X, Xiao G (2007) Tolerance of submergence in rice: gene studies using differential display technique. Chin J Agric Biotechnol 4:139–144
Clarke HJ, Wilson JG, Kuo I, Lulsdorf MM, Mallikarjuna N, Kuo J, Siddique KHM (2006) Embryo rescue and plant regeneration in vitro of selfed chickpea (*Cicer arietinum* L.) and its wild annual relatives. Plant Cell Tissue Organ Cult 85:197–204
Colebatch G, Trevaskis B, Udvardi M (2002) Functional genomics: tools of the trade. New Phytol 153:27–36
Coughlan SJ, Agrawal V, Meyers B (2004) A comparison of global gene expression measurement technologies in *Arabidopsis thaliana*. Comp Funct Genomics 5:245–252
Darwish SA, Pan L, Ide C, Bede JC (2008) Caterpillar-specific gene expression in the legume, *Medicago truncatula*. Plant Mol Biol Rep 26:12–31
Davey MR, Anthony P, Power JB, Lowe KC (2005) Plant protoplasts: status and biotechnological perspectives. Biotechnol Adv 23:131–171
Degenkolbe T, Do PT, Zuther E, Repsilber D, Walther D, Hincha DK, Kohl KI (2009) Expression profiling of rice cultivars differing in their tolerance to long-term drought stress. Plant Mol Biol 69:133–153
Drew RA, Siar SV, O'Brien CM, Magdalita PM, Sajise AGC (2006) Breeding for papaya ringspot virus resistance in *Carica papaya* via hybridisation with *Vasconcellea quercifolia*. Aust J Exp Agric 46:413–418
Dziadczyk P, Bolibok H, Tyrka M, Hortynski JA (2003) In vitro selection of strawberry (*Fragaria × ananassa Duch.*) clones tolerant to salt stress. Euphytica 132:49–55
Ernst D, Liegl I, Kiefer E, Seidlitz HK, Heller W, Sandermann H (2001) Early changes in mRNA populations in leaves of ultraviolet-B-treated European beech (*Fagus sylvatica* L.) seedlings. Acta Physiologiae Plantarum 23:343–349
Escorial MC, Sixto H, GarciaBaudin JM, Chueca MC (1996) In vitro culture selection increases glyphosate tolerance in barley. Plant Cell Tissue Organ Cult 46:179–186
FAO/IAEA. Mutant varieties database, Joint FAO/IAEA Home. http://www-mvd.iaea.org. 2006. 5-1-2009. Ref Type: Online Source
Feng YY, Takahashi H, Akagi H, Mori K (2006) New interspecific rice genome constructions obtained by symmetric cell fusion. Plant Breeding 125:403–404
Fernandez P, Rienzo J, Fernandez L, Hopp H, Paniego N, Heinz R (2008) Transcriptomic identification of candidate genes involved in sunflower responses to chilling and salt stresses based on cDNA microarray analysis. BMC Plant Biol 8. doi:10.1186/1471-2229-8-11
Filipecki M, Malepszy S (2006) Unintended consequences of plant transformation: a molecular insight. J Appl Genet 47:277–286
Finer JJ, Vain P, Jones MW, McMullen MD (1992) Development of the particle inflow gun for DNA delivery to plant cells. Plant Cell Rep 11:323–328
Fraley RT, Horsch RB (1983) In vitro transformation of *Petunia* protoplasts by *Agrobacterium tumefaciens*. J Cell Biochem Supplement 250
Fratini R, Ruiz ML (2006) Interspecific hybridization in the genus *Lens* applying in vitro embryo rescue. Euphytica 150:271–280
Fratini R, Ruiz ML (2008) Micropropagation of intra and interspecific *Lens* hybrids. Euphytica 159:199–206

Fulton RW (1986) Practices and precautions in the use of cross protection for plant virus disease control. Annu Rev Phytopathol 24:67–81

Fusco N, Micheletto L, Dal Corso G, Borgato L, Furini A (2005) Identification of cadmium-regulated genes by cDNA-AFLP in the heavy metal accumulator *Brassica juncea* L. J Exp Bot 56:3017–3027

Gaj MD (2004) Factors influencing somatic embryogenesis induction and plant regeneration with particular reference to *Arabidopsis thaliana* (L.) Heynh. Plant Growth Reg 43:27–47

Ganeshan S, Caswell KL, Kartha KK and Chibbar RN. (2002). Shoot regeneration and proliferation. In: Khachatourians GG, McHughen A, Scorza R, Nip W-K and Hui YH (eds) Transgenic plants and crops. Marcel Dekker Inc., New York, pp 69–84

Ganeshan S, Baga M, Harvey BL, Rossnagel BG, Scoles GJ, Chibbar RN (2003) Production of multiple shoots from thidiazuron-treated mature embryos and leaf-base/apical meristems of barley (*Hordeum vulgare*). Plant Cell Tissue Organ Cult 73:57–64

Ganeshan S, Chodaparambil SV, Baga M, Fowler DB, Hucl P, Rossnagel BG, Chibbar RN (2006) In vitro regeneration of cereals based on multiple shoot induction from mature embryos in response to thidiazuron. Plant Cell Tissue Organ Cult 85:63–73

Ganeshan S, Vitamvas P, Fowler DB, Chibbar RN (2008) Quantitative expression analysis of selected *COR* genes reveals their differential expression in leaf and crown tissues of wheat (*Triticum aestivum* L.) during an extended low temperature acclimation regimen. J Exp Bot 59:2393–2402

Ganeshan S, Denesik T, Fowler DB, Chibbar RN (2009) Quantitative expression analysis of selected low temperature-induced genes in autumn-seeded wheat (*Triticum aestivum* L.) reflects changes in soil temperature. Environ Exp Bot 66:46–53

Ge TM, Lin XH, Qin FL, Yu SW, Yu YJ (2006) Protoplast electrofusion between common wheat (*Triticum aestivum* L.) and Italian ryegrass (*Lolium multiflorum* Lam.) and regeneration of mature cybrids. In Vitro Cell Dev Biol Plant 42:179–187

Goormachtig S, ValerioLepiniec M, Szczyglowski K, VanMontagu M, Holsters M, deBruijn FJ (1995) Use of differential display to identify novel *Sesbania rostrata* genes enhanced by *Azorhizobium caulinodans* infection. Mol Plant Microbe Interact 8:816–824

Green CD, Simons JF, Taillon BE, Lewin DA (2001) Open systems: panoramic views of gene expression. J Immunol Methods 250:67–79

Grosser JW, Ollitrault P, Olivares-Fuster O (2000) Somatic hybridization in citrus: an effective tool to facilitate variety improvement. In Vitro Cell Dev Biol Plant 36:434–449

Grumet R, Sanford JC, Johnston SA (1987) Pathogen-derived resistance to viral infection using a negative regulatory molecule. Virology 161:561–569

Guha S, Maheshwari SC (1964) In vitro production of embryos from anthers of *Datura*. Nature 204:497

Guha S, Maheshwari SC (1966) Cell division and differentiation of embryos in pollen grains of *Datura* in vitro. Nature 212:97

Gurel E, Gurel S, Lemaux PG (2008) Biotechnology applications for sugar beet. Crit Rev Plant Sci 27:108–140

Hamada H, Matsumura H, Tomita R, Terauchi R, Suzuki K, Kobayashi K (2008) SuperSAGE revealed different classes of early resistance response genes in *Capsicum chinense* plants harboring L-3-resistance gene infected with Pepper mild mottle virus. J Gen Plant Pathol 74:313–321

Hamrouni L, Ben Abdallah F, Abdelly C, Ghorbel A (2008) In vitro culture: a simple and efficient way for salt-tolerant grapevine genotype selection. Comptes Rendus Biologies 331:152–163

Hancock JF. 2004. Plant evolution and the origin of crop species. CABI, Wallingford, Oxon, UK

He SZ, Han YF, Wang YP, Zhai H, Liu QC (2009) In vitro selection and identification of sweetpotato (*Ipomoea batatas* (L.) Lam.) plants tolerant to NaCl. Plant Cell Tissue Organ Cult 96:69–74

Hermsen JGT, Verdenius J (1973) Selection from *Solanum tuberosum* group Phureja of genotypes combining high frequency haploid induction with homozygosity for embryo-spot. Euphytica 22:244–259

Hiei Y, Ohta S, Komari T, Kumashiro T (1994) Efficient transformation of rice (*Oryza sativa* L) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. Plant J 6:271–282

Higuchi R, Fockler C, Dollinger G, Watson R (1993) Kinetic PCR analysis - Real-time monitoring of DNA amplification reactions. Bio-Technology 11:1026–1030

Holzberg S, Brosio P, Gross C, Pogue GP (2002) Barley stripe mosaic virus-induced gene silencing in a monocot plant. Plant J 30:315–327

Hooykaas PJJ, Klapwijk PM, Nuti MP, Schilperoort RA, Rorsch A (1977) Transfer of *Agrobacterium tumefaciens* Ti plasmid to avirulent agrobacteria and to *Rhizobium* ex planta. J Gen Microbiol 98:477–484

Horvath H, Jensen LG, Wong OT, Kohl E, Ullrich SE, Cochran J, Kannangara CG, von Wettstein D (2001) Stability of transgene expression, field performance and recombination breeding of transformed barley lines. Theor Appl Genet 102:1–11

Houde M, Diallo A (2008) Identification of genes and pathways associated with aluminum stress and tolerance using transcriptome profiling of wheat near-isogenic lines. BMC Genomics 9:doi:10.1186/1471-2164-9-400

Huang J, Wang MM, Bao YM, Sun SJ, Pan LJ, Zhang HS (2008) SRWD: a novel WD40 protein subfamily regulated by salt stress in rice (*Oryza sativa* L.). Gene 424:71–79

Huetteman CA, Preece JE (1993) Thidiazuron – a potent cytokinin for woody plant tissue culture. Plant Cell Tissue Organ Cult 33:105–119

Ideker T, Galitski T, Hood L (2001) A new approach to decoding life: systems biology. Annu Rev Genomics Hum Genet 2:343–372

Iida A, Morikawa H, Yamada Y (1990) Stable transformation of cultured tobacco cells by DNA-coated gold particles accelerated by gas pressure-driven particle gun. Appl Microbiol Biotechnol 33:560–563

Irian S, Xu P, Dai XB, Zhao PX, Roossinck MJ (2007) Regulation of a virus-induced lethal disease in tomato revealed by LongSAGE analysis. Mol Plant Microbe Interact 20: 1477–1488

Ivashuta S, Imai R, Uchiyama K, Gau M (1999) The coupling of differential display and AFLP approaches for nonradioactive mRNA fingerprinting. Mol Biotechnol 12:137–141

Jain SM (2001) Tissue culture-derived variation in crop improvement. Euphytica 118:153–166

Jansen RC, Nap JP (2001) Genetical genomics: the added value from segregation. Trends Genet 17:388–391

Jayaraman A, Puranik S, Rai NK, Vidapu S, Sahu PP, Lata C, Prasad M (2008) cDNA-AFLP analysis reveals differential gene expression in response to salt stress in foxtail millet (*Setaria italica* L.). Mol Biotechnol 40:241–251

Jung SH, Lee JY, Lee DH (2003) Use of SAGE technology to reveal changes in gene expression in Arabidopsis leaves undergoing cold stress. Plant Mol Biol 52:553–567

Kang MS, Subudhi PK, Baisakh N and Priyadarshan PM. (2007). Crop breeding methodologies: classic and modern. In: Kang MS and Priyadarshan PM (eds) Breeding major food staples. Blackwell, Oxford, UK, pp 5–40

Karadimova M, Djambova G (1993) Increased NaCl-tolerance in wheat (*Triticum aestivum* L. and *T. durum* Desf.) through in vitro selection. In Vitro Cell Dev Biol Plant 29:180–182

Karp A (1995) Somaclonal variation as a tool for crop improvement. Euphytica 85:295–302

Kasha KJ, Kao KN (1970) High frequency haploid production in barley (*Hordeum vulgare* L.). Nature 225:874–876

Kiekens R, Vercauteren A, Moerkerke B, Goetghebeur E, Van Den Daele H, Sterken R, Kuiper M, van Eeuwijk F, Vuylsteke M (2006) Genome-wide screening for cis-regulatory variation using a classical diallel crossing scheme. Nucl Acids Res 34:3677–3686

Kiiskinen M, Korhonen M, Kangasjarvi J (1997) Isolation and characterization of cDNA for a plant mitochondrial phosphate translocator (*Mpt1*): ozone stress induces *Mpt1* mRNA accumulation in birch (*Betula pendula* Roth). Plant Mol Biol 35:271–279

Kinoshita T, Mori K (1991) Somaclonal selection of physiological mutants through rice cell culture. Cereal Res Commun 19:131–164

Klein DT, Klein RM (1953) Transmittance of tumor-inducing ability to avirulent crown gall and related bacteria. J Bacteriol 66:220–228

Klein TM, Fromm M, Weissinger A, Tomes D, Schaaf S, Sletten M, Sanford JC (1988a) Transfer of foreign genes into intact maize cells with high velocity microprojectiles. Proc Natl Acad Sci USA 85:4305–4309

Klein TM, Harper EC, Svab Z, Sanford JC, Fromm ME, Maliga P (1988b) Stable genetic transformation of intact *Nicotiana* cells by the particle bombardment process. Proc Natl Acad Sci USA 85:8502–8505

Kloosterman B, Vorst O, Hall RD, Visser RGF, Bachem CW (2005) Tuber on a chip: differential gene expression during potato tuber development. Plant Biotechnol J 3:505–519

Krikorian AD, Berquam DL (1969) Plant cell and tissue cultures – role of Haberlandt. Bot Rev 35:59–67

Krishnaswamy S, Sanjeeva S, Mohammadi M, Rahman M, Deyholos M, Kav N (2008) Transcriptional profiling of pea ABR17 mediated changes in gene expression in *Arabidopsis thaliana*. BMC Plant Biol 8. doi:10.1186/1471-2229-8-91

Kuhn E (2001) From library screening to microarray technology: strategies to determine gene expression profiles and to identify differentially regulated genes in plants. Ann Bot 87:139–155

Kumagai MH, Donson J, Dellacioppa G, Harvey D, Hanley K, Grill LK (1995) Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived RNA. Proc Natl Acad Sci USA 92:1679–1683

Kwiatkowska D (2008) Flowering and apical meristem growth dynamics. J Exp Bot 59:187–201

Lang P, Zhang CK, Ebel RC, Dane F, Dozier WA (2005) Identification of cold acclimated genes in leaves of *Citrus unshiu* by mRNA differential display. Gene 359:111–118

Lao M, Arencibia AD, Carmona ER, Acevedo R, Rodriguez E, Leon O, Santana I (2008) Differential expression analysis by cDNA-AFLP of *Saccharum* spp. after inoculation with the host pathogen *Sporisorium scitamineum*. Plant Cell Rep 27:1103–1111

Larkin PJ, Scowcroft WR (1981) Somaclonal variation – A novel source of variability from cell cultures for plant improvement. Theor Appl Genet 60:197–214

Lefrancois C, Chupeau Y, Bourgin JP (1993) Sexual and somatic hybridization in the genus *Lycopersicon*. Theor Appl Genet 86:533–546

Li H, Zhang P, Zha X, Xia X, He Z (2007) Isolation of differentially expressed genes from wheat cultivars Jinan 17 and Yumai 34 with good bread quality under heat stress during grain filling stage. Acta Agronomica Sinica 33:1644–1653

Liang P, Pardee AB (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 257:967–971

Liu H, Zhang H, Tan Z, Huang Y (2007) Isolation and expression of genes involved in tolerance. J Maize Sci 15:26–30

Liu JH, Xu XY, Deng XX (2005) Intergeneric somatic hybridization and its application to crop genetic improvement. Plant Cell Tissue Organ Cult 82:19–44

Lockhart DJ, Dong HL, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang CW, Kobayashi M, Horton H, Brown EL (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol 14:1675–1680

Long TA, Rady SM, Benfey PN (2008) Systems approaches to identifying gene regulatory networks in plants. Annu Rev Cell Dev Biol 24:81–103

Lu C-Y (1993) The use of thidiazuron in tissue culture. In Vitro Cellular Dev Biol Plant 29P:92–96

Lu M, Li X, Chen J, Chen L, Qian C (2005) Study on chilling tolerance of cucumber during germination and expression of lysine decarboxylase gene. Scientia Agricultura Sinica 38:2492–2495

Luhs W, Friedt W (1994) Present state and prospects of breeding rapeseed (*Brassica napus*) with a maximum erucic acid content for industrial applications. Fett Wissenschaft Technologie – Fat Sci Technol 96:137–146

Lynch PT, Davey MR, Power JB (1993) Plant protoplast fusion and somatic hybridization. Methods Enzymol 221:379–393

Malatrasi M, Close TJ, Marmiroli N (2002) Identification and mapping of a putative stress response regulator gene in barley. Plant Mol Biol 50:143–152

Mao CZ, Yi K, Yang L, Zheng BS, Wu YR, Liu FY, Wu P (2004) Identification of aluminium-regulated genes by cDNA-AFLP in rice (*Oryza sativa* L.): aluminium-regulated genes for the metabolism of cell wall components. J Exp Bot 55:137–143

Maqbool A, Zahur M, Irfan M, Younas M, Barozai K, Rashid B, Husnain T, Riazuddin S (2008) Identification and expression of six drought-responsive transcripts through differential display in desi cotton (*Gossypium arboreum*). Mol Biol 42:492–498

Marcotrigiano M (1997) Chimeras and variegation: Patterns of deceit. Hortscience 32:773–784

McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeting induced local lesions in genomes (TILLING) for plant functional genomics. Plant Physiol 123:439

Mckinney HH (1927) Factors affecting certain properties of a mosaic virus. J Agric Res 35:1–12

Mckinney HH (1929) Mosaic diseases in the Canary Islands, west Africa, and Gibraltar. J Agric Res 39:557–578

Mckinney HH (1937) Virus mutation and the gene concept. J Hered 28:51–57

Millam S, Payne LA, Mackay GR (1995) The integration of protoplast fusion-derived material into a potato breeding program – A review of progress and problems. Euphytica 85:451–455

Morikawa H, Iida A, Yamada Y (1989) Transient expression of foreign genes in plant cells and tissues obtained by a simple biolistic device (particle gun). Appl Microbiol Biotechnol 31:320–322

Muangprom A, Mauriera I, Osborn TC (2006) Transfer of a dwarf gene from *Brassica rapa* to oilseed *B. napus*, effects on agronomic traits, and development of a 'perfect' marker for selection. Mol Breed 17:101–110

Muller HJ (1927) Artificial transmutation of the gene. Science 66:84–87

Muller HJ (1928) The production of mutations by x-rays. Proc Natl Acad Sci USA 14:714–726

Mullis KB, Faloona FA (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol 155:335–350

Murthy BNS, Murch SJ, Saxena PK (1998) Thidiazuron: a potent regulator of in vitro plant morphogenesis. In Vitro Cell Dev Biol Plant 34:267–275

Nei M (1963) Efficiency of haploid method of plant breeding. Heredity 18:95–100

Oard JH, Paige DF, Simmonds JA, Gradziel TM (1990) Transient gene expression in maize, rice and wheat cells using an airgun apparatus. Plant Physiol 92:334–339

Oerke E-C, Dehne H-W, Schönbeck F and Weber A. 1994. Crop production and crop protection: estimated losses in major food and cash crops. Elsevier, Amsterdam.

Orczyk W, Przetakiewicz J, Nadolska-Orczyk A (2003) Somatic hybrids of *Solanum tuberosum* – application to genetics and breeding. Plant Cell Tissue Organ Cult 74:1–13

Petolino JF, Hopkins NL, Kosegi BD, Skokut M (2000) Whisker-mediated transformation of embryogenic callus of maize. Plant Cell Rep 19:781–786

Phillips RL, Kaeppler SM, Olhoft P (1994) Genetic instability of plant tissue cultures – Breakdown of normal controls. Proc Natl Acad Sci USA 91:5222–5226

Pofelis S, Le H, Grant WF (1992) The development of sulfonylurea herbicide-resistant birdsfoot trefoil (*Lotus corniculatus*) plants from in vitro selection. Theor Appl Genet 83:480–488

Potokina E, Caspers M, Prasad M, Kota R, Zhang H, Sreenivasulu N, Wang M, Graner A (2004) Functional association between malting quality trait components and cDNA array based expression patterns in barley (*Hordeum vulgare* L.). Mol Breed 14:153–170

Queiros F, Fidalgo F, Santos I, Salema R (2007) In vitro selection of salt tolerant cell lines in *Solanum tuberosum* L. Biol Plant 51:728–734

Raghavan V (1985) The applications of embryo rescue in agriculture. International Rice Research Institute Biotechnology in International Agricultural Research; Inter-Center Seminar on International Agricultural Research Centers and Biotechnology, Manila, Philippines, Apr 23–27, 1984 Viii+435P International Rice Research Institute pp 189–198

Raghavan V (1986) Embryogenesis in angiosperms. A developmental and experimental study. Cambridge University Press, Cambridge, UK

Rajesh P, Gupta V, Ranjekar P, Muehlbauer F (2003) Functional genome analysis using DDRT for ascochyta blight resistance in chickpea. International Chickpea and Pigeonpea Newsletter 10:35–37

Rani V, Raina SN (2000) Genetic fidelity of organized meristem-derived micropropagated plants: A critical reappraisal. In Vitro Cell Dev Biol Plant 36:319–330

Ratcliff F, Harrison BD, Baulcombe DC (1997) A similarity between viral defense and gene silencing in plants. Science 276:1558–1560

Reinartz J, Bruyns E, Lin JZ, Burcham T, Brenner S, Bowen B, Kramer M, Woychik R (2002) Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. Brief Funct Genomics Proteomics 1:95–104

Robinson SJ, Parkin IAP (2008) Differential SAGE analysis in Arabidopsis uncovers increased transcriptome complexity in response to low temperature. BMC Genomics 9. doi:10.1186/1471-2164-9-434

Rodrangboon P, Pongtongkam P, Suputtitada S, Adachi T (2002) Abnormal embryo development and efficient embryo rescue in interspecific hybrids, *Oryza sativa* × *O-minuta* and *O-sativa* × *O-officinalis*. Breed Sci 52:123–129

Rout GR, Mohapatra A, Jain SM (2006) Tissue culture of ornamental pot plant: A critical review on present scenario and future prospects. Biotechnol Adv 24:531–560

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Sanford JC (1990) Biolistic plant transformation. Physiologia Plantarum 79:206–209

Sanford JC, Johnston SA (1985) The concept of parasite-derived resistance – Deriving desistance genes from the parasites own genome. J Theor Biol 113:395–405

Sarita C, Yu K (2006) Gene profiling of plants with cDNA-AFLP. Floriculture, ornamental and plant biotechnology 489–497

Sarrafi A, Amrani N, Alibert G (1994) Haploid regeneration from tetraploid wheat using maize pollen. Genome 37:176–178

Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH (2003) Genetics of gene expression surveyed in maize, mouse and man. Nature 422:297–302

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470

Schenk PM, Kazan K, Wilson I, Anderson JP, Richmond T, Somerville SC, Manners JM (2000) Coordinated plant defense responses in Arabidopsis revealed by microarray analysis. Proc Natl Acad Sci USA 97:11655–11660

Sharma DR, Kaur R, Kumar K (1996) Embryo rescue in plants – A review. Euphytica 89:325–337

Shi C, Uzarowska A, Ouzunova M, Landbeck M, Wenzel G, Lubberstedt T (2007) Identification of candidate genes associated with cell wall digestibility and eQTL (expression quantitative trait loci) analysis in a *Flint* × *Flint* maize recombinant inbred line population. BMC Genomics 8

Shrawat AK, Lorz H (2006) Agrobacterium-mediated transformation of cereals: a promising approach crossing barriers. Plant Biotechnol J 4:575–603

Simoes-Araujo JL, Alves-Ferreira M, Rumjanek NG, Margis-Pinheiro M (2008) VuNIP1 (NOD26-like) and VuHSP17.7 gene expression are regulated in response to heat stress in cowpea nodule. Environ Exp Bot 63:256–265

Sleper DA and Poehlman JM. (2006). Plant breeders and their work. Breeding field crops, 5th edn. Blackwell, Ames, Iowa, USA, pp 3–16

Stadler LJ (1928) Mutations in barley induced by X-rays and radium. Science 68:186–187

Steiner B, Schieszl K, Litwicka EP, Kurz H, Lemmens M, Jia H, Muehlbauer G, Buerstmayr H (2008) Gene expression analysis of related wheat lines with contrasting levels of head blight resistance after *Fusarium graminearum* inoculation. Cereal Res Commun 36: 267–269

Strauss A, Gebhardt C and King PJ. (1980). Methods for selection of drug-resistant plant cell cultures. In: Ingram DS, Helgeson JP (eds) Tissue culture methods for plant pathologists. Blackwell, Oxford, UK, pp 239–248

Svabova L, Lebeda A (2005) In vitro selection for improved plant resistance to toxin-producing pathogens. J Phytopathol 153:52–64

Szymkowiak EJ, Sussex IM (1996) What chimeras can tell us about plant development. Annu Rev Plant Physiol Plant Mol Biol 47:351–376

Taregyan MR, Mortimer AM, Putwain PD, Collin HA (2001) Selection for resistance to the herbicide imazethapyr in somaclones of soyabean. Weed Res 41:143–154

Thieme R, Rakosy-Tican E, Gavrilenko T, Antonova O, Schubert J, Nachtigall M, Heimbach U, Thieme T (2008) Novel somatic hybrids (*Solanum tuberosum* L. plus *Solanum tarnii*) and their fertile BC1 progenies express extreme resistance to potato virus Y and late blight. Theor Appl Genet 116:691–700

Tian LL, Wang YJ (2008) Seedless grape breeding for disease resistance by using embryo rescue. Vitis 47:15–19

Tikhenko N, Rutten T, Voylokov A, Houben A (2008) Analysis of hybrid lethality in F-1 wheat-rye hybrid embryos. Euphytica 159:367–375

Tooke F, Battey N (2003) Models of shoot apical meristem function. New Phytol 159:37–52

Torney F, Trewyn BG, Lin VSY, Wang K (2007) Mesoporous silica nanoparticles deliver DNA and chemicals into plants. Nat Nanotechnol 2:295–300

Touraev A, Forster BP and Jain SM. 2009. Advances in haploid production in higher plants. Springer, Heidelberg, Germany

Unver T, Bozkurt O, Akkaya MS (2008) Identification of differentially expressed transcripts from leaves of the boron tolerant plant *Gypsophila perfoliata* L. Plant Cell Rep 27:1411–1422

van Harten AM. 1998. Mutation breeding:theory and practical applications. Cambridge University Press, Cambridge, UK

van Kammen A (1997) Virus-induced gene silencing in infected and transgenic plants. Trends Plant Sci 2:409–411

Vazquez AM (2001) Insight into somaclonal variation. Plant Biosyst 135:57–62

Veen R, Dulk-Ras H, Bisseling T, Schilperoort R, Hooykaas P (1988) Crown gall tumor and root nodule formation by the bacterium *Phyllobacterium myrsinacearum* after the introduction of an Agrobacterium Ti plasmid or a Rhizobium Sym plasmid. Mol Plant Microbe Interact 1:231–234

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science 270:484–487

Ventelon-Debout M, Tranchant-Dubreuil C, Nguyen TTH, Bangratz M, Sire C, Delseny M, Brugidou C (2008) Rice Yellow Mottle Virus stress responsive genes from susceptible and tolerant rice genotypes. BMC Plant Biol 8. doi:10.1186/1471-2229-8-26

Veres G, Gibbs RA, Scherer SE, Caskey CT (1987) The molecular basis of the sparse fur mouse mutation. Science 237:415–417

Visioli G, Maestri E, Marmiroli N (1997) Differential display-mediated isolation of a genomic sequence for a putative mitochondrial LMW HSP specifically expressed in condition of induced thermotolerance in *Arabidopsis thaliana* (L) *Heynh*. Plant Mol Biol 34:517–527

Vuylsteke M, van Eeuwijk F (2008) The use of general and specific combining abilities in a context of gene expression relevant to plant breeding. Euphytica 161:115–122

Waara S, Glimelius K (1995) The potential of somatic hybridization in crop breeding. Euphytica 85:217–233

Wagner A (2002) Estimating coarse gene network structure from large-scale gene perturbation data. Genome Res 12:309–315

Walter S, Brennan JM, Arunachalam C, Ansari KI, Hu XJ, Khan MR, Trognitz F, Trognitz B, Leonard G, Egan D, Doohan FM (2008) Components of the gene network associated with genotype-dependent response of wheat to the *Fusarium* mycotoxin deoxynivalenol. Funct Integr Genomics 8:421–427

Wang JPP, Bughrara SS (2007) Monitoring of gene expression profiles and identification of candidate genes involved in drought responses in *Festuca mairei*. Mol Genet Genomics 277:571–587

Wang YP, Sonntag K, Rudloff E (2003) Development of rapeseed with high erucic acid content by asymmetric somatic hybridization between *Brassica napus* and *Crambe abyssinica*. Theor Appl Genet 106:1147–1155

Wang YP, Sonntag K, Rudloff E, Chen JM (2005) Intergeneric somatic hybridization between *Brassica napus* L. and *Sinapis alba* L. J Integr Plant Biol 47:84–91

Wang YP, Sonntag K, Rudloff E, Groeneveld I, Gramenz J, Chu CC (2006) Production and characterization of somatic hybrids between *Brassica napus* and *Raphanus sativus*. Plant Cell Tissue Organ Cult 86:279–283

Weiberg A, Pohler D, Morgenstern B, Karlovsky P (2008) Improved coverage of cDNA-AFLP by sequential digestion of immobilized cDNA. BMC Genomics 9. doi:10.1186/1471-2164-9-480

Weissinger A, Tomes D, Sanford J, Kline T, Fromm M (1987) Microprojectile bombardment for maize transformation. In Vitro Cell Dev Biol 23:A75

West MAL, Kim K, Kliebenstein DJ, van Leeuwen H, Michelmore RW, Doerge RW, Clair DAS (2007) Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. Genetics 175:1441–1450

Winicov I (1996) Characterization of rice (*Oryza sativa* L) plants regenerated from salt-tolerant cell lines. Plant Sci 113:105–111

Yamakawa H, Hirose T, Kuroda M, Yamaguchi T (2007) Comprehensive expression profiling of rice grain filling-related genes under high temperature using DNA microarray. Plant Physiol 144:258–277

Yang DL, Li W, Li S, Yang XL, Wu JL, Cao ZY (2007) In vitro embryo rescue culture of F-1 progenies from crosses between diploid and tetraploid grape varieties. Plant Growth Reg 51:63–71

Yih TC, Al-Fandi M (2006) Engineered nanoparticles as precise drug delivery systems. J Cell Biochem 97:1184–1190

Yu CY, Chun SC, Lim JD, Kim SH, Kang WH, Chung IM (2004) Tetrapyrrole accumulation in in vitro selected somaclones of acifluorfen-tolerant *Solanum ptycanthum* Dun. Plant Cell Tissue Organ Cult 76:167–174

Yu X, Lu X, Chen Y, Li d, Yao J, Zhang H (2007) Differential gene expression of *Senecio* × *hybridus* under heat stress. J Agric Biotechnol 15:459–463

Zhang BH, Liu F, Wang QL, Zhang WS (2001) Selection for salt tolerance in cotton tissue culture and plant regeneration from NaCl-tolerant embryogenic callus. Isr J Plant Sci 49:187–191

Zhang C, Chen SY (1996) Analysis of genes specifically expressed under salt stress in salt-tolerant mutant of rice by using DDRT-PCR technique. Sci Chin Ser C-Life Sci 39:385–394

Zhang JJ, He ZH, Tian H, Zhu GH, Peng XX (2007) Identification of aluminium-responsive genes in rice cultivars with different aluminium sensitivities. J Exp Bot 58:2269–2278

Zhu JW, Xu YP, Zhang ZX, Cao WY, Cai XZ (2008) Transcript profiling for Avr4/Cf-4- and Avr9/Cf-9-dependent defence gene expression. Eur J Plant Pathol 122:307–314

Zhu YS, Chen BT, Yu SW, Zhang DP, Zhang XQ, Yan QS (2004) Transfer of bacterial blight resistance from *Oryza meyeriana* to *O. sativa* L. by asymmetric somatic hybridization. Chin Sci Bull 49:1481–1484

Zhuang YL, Ren GJ, Zhu Y, Hou GH, Qu X, Li ZX, Yue GD, Zhang JR (2008) Transcriptional profiles of immature ears and tassels in maize at early stage of water stress. Biologia Plantarum 52:754–758

Zwiegelaar M, Dubery IA (2006) Early activation of cell wall strengthening-related gene transcription in cotton by a *Verticillium dahliae* elicitor. S Afr J Bot 72:467–472

Chapter 4
Bioinformatics Tools for Crop Research and Breeding

Jayashree B and Dave Hoisington

Abstract Crop improvement has always been, but will be even more so in the twenty-first century, an information intensive process. For effective and efficient improvement, a range of activities from molecular biology to genetics to indirect selection must now be involved. The rate of progress made by any breeding programme depends as much on the efficient integration of information from these activities as it does on the activities themselves. Plant breeders are now realizing the importance of innovative approaches that include the use of a range of molecular methods and their outputs, and the possibilities of transferring this information from model species to cultivated crops. The use of these high throughput methods in model crops has already generated a large amount of public resources such as databases containing genetic resource, genomic and genetic information; tools for the effective analysis, data mining and visualization of such information; and semantic web resources for data integration. In this chapter, we highlight the role and contributions of bioinformatics to crop research and breeding by focusing on the bioinformatics resources that are available for crop science research and breeding, and indicating gaps that need to be bridged that will allow scientists to access, transfer and integrate data with ease.

4.1 Introduction

The growing world-wide demand for food is placing increasing pressure on crop breeding programs to produce cultivars that can adapt to a range of environments without compromising on quality and yield. As such, crop breeding efforts focus on developing new varieties with improved resistance to fungal, insect or viral diseases, tolerance to abiotic stresses such as drought, cold, salt, dehydration, heavy metal toxicity and numerous quality attributes such as taste, size, shape, color and ease of cooking. In addition there is a growing need to provide for nutritional deficiencies, especially in the developing countries

B. Jayashree and D. Hoisington (✉)
International Crops Research Institute for the Semi Arid Tropics, Patancheru, 502 324, Andhra Pradesh, India
e-mail: d.hoisington@CGIAR.ORG

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_4,

through biofortified crops. With an ever-increasing number of desirable traits that must be integrated into a cultivar, crop improvement programs are at an interesting juncture. The combination of existing knowledge and resources with modern structural and functional genomics provides the opportunity to study the genetic, biochemical and physiological basis of complex traits. Efforts currently center on capturing information from model and better-studied crops in order to define genes for important traits. With the advent of high throughput technologies a number of initiatives have emerged, for example large scale mapping studies, genome-wide expression studies and high throughput screening of genotypes and phenotypes along with corresponding bioinformatics resources. It is now becoming clear that crop improvement programs will benefit hugely from a judicious use of these resources coupled with crop genetic resources, which are the basic materials for breeding programs. Genetic resource collections available to breeders are being characterized for diversity so breeders can have access to 'core' collections that contain as much genetic variability as possible. The advances in plant genetics and genomics offer opportunities so far unavailable, for discovering the function of genes and the potential to manipulate them for crop improvement. With so much information being produced that could be of potential use to the breeder, the difficulty is in making sense of all the data so as to facilitate knowledge driven crop selection. Bioinformatics is emerging as the glue that brings these different kinds of data together; as a discipline it spans the realm from scientific software development to meaningful knowledge discovery. A review of current bioinformatics resources, tools and methods available for the purpose of crop improvement gives us an idea of ground covered so far and what is desirable to achieve in the coming years.

4.2 Bioinformatics Resources Available for Crop Research

The burgeoning information from genomics is due to innovative technologies like DNA microarrays, high throughput genotyping and Next Generation Sequencing. Most modern data generation projects have seen a concomitant development of databases to store, access and query data. These data resources are usually made available through the web, store varying kinds of information and are available at different locations. The very distributed nature of this information throws up interesting challenges – that of interoperability of databases that will allow data integration, the use or lack of common vocabularies that will allow comparison of the data and the varying levels of data annotation and curation available that reflects on data quality. Databases can no longer be passive storehouses of information, they need to link to various types of data to be useful.

4.2.1 Data Resources

There are a considerable number of quality databases devoted to crops that allow access to users through GUIs (Graphical User Interfaces). Amongst the online resources listing key databases of value is the *Nucleic Acids Research* online Molecular Biology database collection (http://www.oxfordjournals.org/nar/database/a/).

The number of databases in this collection is 1,170 as of January 2009, with 78 plant specific databases. The collection lists high quality, comprehensive databases with value added in the form of manual curation. The bioinformatics links directory (http://bioinformatics.ca/links directory) is an actively maintained compilation of servers hosting bioinformatics databases with features for improved navigation and accessibility.

Table 4.1 lists popular as well as lesser known crop species and multi-crop species databases covering genotype, phenotype, taxonomy and genomic information. Besides

Table 4.1 Species and clade specific crop databases

Database	Species	Primary site	Database contents
BeanGenes	Phaseolus and Vigna	http://beangenes.cws.ndsu.nodak.edu/	Genetic, germplasm, phenotypic and pathology data
CR-EST (crop EST)	Barley, pea, potato, petunia, tobacco, wheat	http://pgrc.ipk-gatersleben.de/est/index.php	Genomic data
FoggDB	Forage grasses	http://www.igergru.bbsrc.ac.uk/Welcome/IGER/foggdb/foggdb.htm	Genomic data
GDR (genome database for Rosaceae)	Apple, pear, prunus, raspberry, strawberry, prunus	http://www.bioinfo.wsu.edu/gdr/	Genomic data
Graingenes	Wheat, rye, barley, oats, sugarcane and relatives	http://wheat.pw.usda.gov/GG2/index.shtml	Genetic, genomic, expression, phenotypic and taxonomy data
Gramene	Rice, Sorghum, maize, wheat, rye, millets, *Arabidopsis*	http://www.gramene.org/	Genetic, genomic and pathway data
JCVI (TIGR)	25 crops	http://www.tigr.org/	Genomic data
LIS (Legume Information Service)	17 legume species	http://www.comparative-legumes.org/	Genetic and genomic data
MaizeGDB	Maize	http:/www.maizegdb.org	Genetic, genomic and phenotypic data
MIPSPlantsDB	Multispecies	http://mips.gsf.de/projects/plants	Genomic data
Soybase	Soybean	http://soybase.agron.iastate.edu/	Genetic, genomic and phenotypic data
TAIR (The Arabidopsis Information Resource)	Arabidopsis	http://www.arabidopsis.org/	Genetic, genomic and gene expression data
TIGR plant transcript assemblies (TA) database	Multispecies	http://plantta.tigr.org	EST and cDNA data
PlantGDB	Multispecies	http://www.plantgdb.org/	Genomic data
UKCropNet	Central multispecies database querying system	http://ukcrop.net/db.html	Genetic, genomic and pathway data

the databases listed in this table, highly specialized databases derived from the research on model crops are available on the web. PathoPlant® is a database on plant–pathogen interactions and signal transduction reactions using microarray gene expression data from *Arabidopsis thaliana* subjected to pathogen infection and elicitor treatment (http://www.pathoplant.de). The cereal small RNA DB (CSRDB) consists of large scale datasets of maize and rice smRNA generated by high throughput pyrosequencing, mapped to the rice and maize genomic sequence (http://sundarlab.ucdavis.edu/smrnas/). Resources for comparative genomics include the POGs/Plant RBP (putative orthologous groups/plant RNA binding proteins, http://plantrbp.uoregon.edu/), ATTED-11 (*A. thaliana* trans factor and *cis* element prediction database) with information on function and regulation of particular genes and gene networks (http://www.atted.bio.titech.ac.jp). The GABI-Kat SimpleSearch is an *Arabidopsis* T-DNA mutant database containing >108,000 mapped FSTs (flanking sequence tags) from ~64,000 lines which cover 64% of all annotated *A. thaliana* protein coding genes (http://www.GABI-Kat.de). The plantTFDB stores information on transcription factors predicted from 22 species: 5 model organisms and 17 plants (http://planttfdb.cbi.pku.edu.cn/). PlantQTL-GE is a database system for identifying candidate genes in rice and *Arabidopsis* by gene expression and QTL information. The database includes genes, gene expression information, ESTs and genetic markers from multiple sources (http://www.scbit.org/qtl2gene/new/). The plant promoter database provides promoter annotations in Arabidopsis and rice (http://www.ppdb.gene.nagoya-u.ac.jp). MetaCrop is a database of crop plant metabolism including pathway diagrams, reactions, transport processes and reaction kinetics besides taxonomy and literature (http://metacrop.ipk-gatersleben.de). All the databases referred to here have been published over the period 2006–2009 and show the differences in resources available on model crops as compared to orphan crops.

4.2.2 Web and Web Services

Most bioinformatics databases and analytical services are available through the Internet. The user may need to interact with many of these in concert to extract different kinds of data, and compare, integrate and format data for submission to an analysis program. Web interfaces are not really suited to handle bulk data export from databases and programmatic access to data is needed to retrieve large quantities of data and format it for submission to analytical tools. Thus data source providers have begun to allow multiple modes of data retrieval and view, from HTML (hypertext markup language), XML (extensible markup language), and SQL (structured query language), to SOAP (simple object access protocol, used in web services) besides allowing hook up to third party analysis tools. Markup languages like HTML and XML provide the means to describe the structure of text-based information. XML defines a way to add markup to information as well as assign meanings to data explicitly, thus facilitating machine readability. Where meaning is implicit only a person with knowledge about the data can

understand and interpret it, but where meaning is explicit, the data becomes interpretable by retrieving software. Examples of databases that provide XML access include INSD_v1.4 that provides access to the EMBL/DDBJ/Genbank sequence records in XML, while GrainGenes (http://wheat.pw.usda.gov/cgi-bin/graingenes/sql.cgi) provides SQL access to its database. Web services provide a programmatic interface to databases and web-based tools and are increasingly being used to automate execution of the data retrieval and analysis steps. The users can look up XML-based web service registries that list name, products, locations and services offered by the web service provider on the Internet. Examples of popular bioinformatics web services projects include BioMoby (The BioMoby Consortium 2008) and myGrid (www.mygrid.org.uk). The web service registry here is different from traditional web services in that it uses the meaning of terms in the biological vocabulary (semantics) to mediate web service discovery and invocation. This helps overcome the problem inherent to biological data – that of inconsistent data type. The Virtual Plant Information network hosted at the NCGR is another network of data and service providers based on the semantic web services platform (http://vpin.ncgr.org/). This network differs from paradigmatic web services in that it does not use SOAP for information exchange but instead relies on http and the web ontology language (OWL-DL), a web standard for information processing. VPIN has a web front end that allows users to find disparate data and services based on lexical and semantic criteria. The DAS (Distributed Annotation System) is another data retrieval protocol that can be used for the exchange of biological sequence annotation. It allows a single machine to gather up sequence annotation information from multiple distant web sites, collate the information, and display it to the user in a single view (Prlic et al. 2007). A small number of plant/crop data sources are now beginning to make their data available through web services.

4.2.3 Data Integration and the Semantic Web

The bioinformatics community has been experimenting with two methods of biological database integration. In the data warehouse approach data from different data sources is translated into a local warehouse and all queries are executed on the warehouse. Examples include DataFoundry (Critchlow et al. 2000) and BioWarehouse (Lee et al. 2006). The warehouse needs to be updated frequently to reflect the modifications in the source databases. The second method is the federated database approach, where the query is executed on a single federated schema that is an integration of component database schemas (a schema can be considered to be the layout of a database). A good example of a federated query system designed specifically for use with large datasets is BioMart (http://www.ebi.ac.uk/biomart). Major databases that implement BioMart include Ensembl, a software system that produces and maintains automatic annotation on selected eukaryotic genomes (http://www.ensembl.org/index.html); VEGA (http://vega.sanger.ac.uk/

index.html), the manually annotated Vertebrate Genome Annotation; dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/), and the Single Nucleotide Polymorphism database of NCBI.

At the level of data integration, most methods followed so far are based on syntax; explicit cross references and common contents which heavily rely on manual annotation of data that can be time consuming, error prone and expensive. Several bioinformatics databases are now moving towards a standardized method of describing their data so that data retrieval and integration can be independent of source database schemas. In the semantic web approach to data integration, the web is no longer a network of documents but a network of data and knowledge. The semantic web provides common formats and languages for consistent and standardized data representation and exchange. In the context of databases, it means that data will be encoded with additional meta-information that will provide context to the data which is made available through web services. That encoding makes use of ontologies. The key role of ontologies with respect to database systems is to specify a data modeling representation at a level of abstraction above specific database designs (logical or physical). Due to their independence from lower level data models, ontologies can be used for integrating heterogeneous databases, enabling interoperability among disparate systems, and specifying interfaces that can be queried independently.

Ontologies are part of the Semantic Web architecture (see Fig. 4.1, the W3C or World-Wide Web consortium develops common protocols for the World-Wide Web that promotes its evolution and interoperability). Ontologies define a set of representational classes, attributes and relationships with which to model a domain of knowledge. Take for example the Gene Ontology (GO), a community effort to provide controlled vocabulary to describe gene and gene product attributes in any organism. When one database describes a piece of data as being "a gene as defined by the Gene Ontology", the data consumer can use or not use the data based on the understanding of "a gene as defined by the Gene Ontology" rather than worry about datasource specific definitions of the 'gene'. Similarly, the Plant Ontology Consortium (POC) (www.plantontology.org) is a collaborative effort to develop simple yet robust and extensible controlled vocabularies that accurately reflect the biology of plant structures and developmental stages. There is Trait Ontology (TO) for traits and phenotype data (http://www.gramene.org/plant_ontology/trait.ontology). MyGrid and BioMoby ontologies are for the semantic discovery of bioinformatics services. They use ontological reasoning over both data type and service definitions for service discovery. Clients can interact with multiple sources of biological data, regardless of the underlying database format/schema. While ontologies are being implemented only by a small number of data sources, they become relevant to the interoperability of expanding database collections (http://www.gramene.org/resources/plant_databases.pdf). There are published examples to show the application of semantic web technologies to build data warehouses that facilitate integration of genomic/proteomic data (Smith et al. 2007).

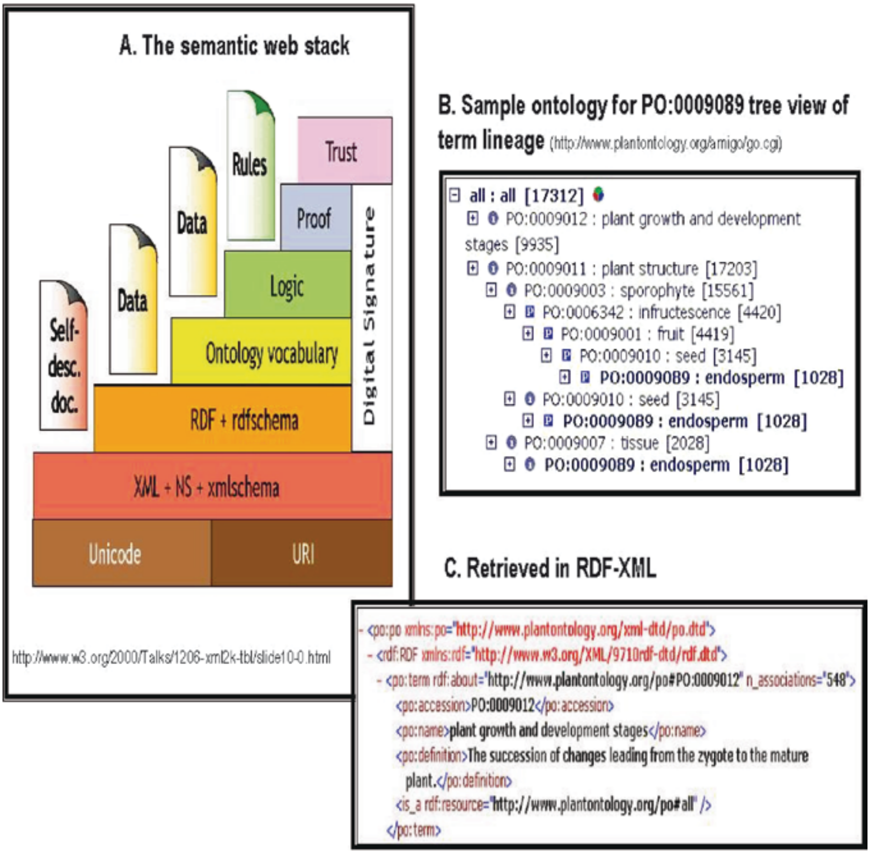


Fig. 4.1 Semantic web for data integration through metadata-based reasoning. (**A**) The semantic web stack (**B**) a sample representation of an ontology for a term derived from the plant ontology consortium website (C) the same ontology retrieved in RDF-XML representation

4.2.4 Bioinformatics Tools for Comparative Genomics

Comparative genomics in silico offers the possibility of linking crops through their sequences and genome maps to provide keys to understanding how genes and genomes are structured, how they function and evolve. Significant synteny amongst the cereal crops has allowed the alignment of major economically important qualitative or quantitative trait loci across specific chromosomal regions. This has facilitated candidate gene and flanking marker identification and their comparisons with annotated sequences from model crops, important for the application of marker-assisted selection. The benefits of transferring genomics information from model to orphan crops could take one of several forms: (a) the identification of potentially

useful variants, (b) Marker Assisted Selection (MAS) of desired alleles and allele combinations, and (c) cloning and direct transfer of desirable alleles among taxa (Nelson et al. 2004).

Very large collections of bioinformatics tools have been developed on the open source model, meaning that they are freely available to use and learn from and improve upon. Of the tools available for comparative genomics, sequence alignment tools are the most commonly used. These tools can be used to query databases for sequences similar to an input sequence, find previously characterized sequences, detect relationships amongst sequences, as well as identify possible functions based on similarity to known sequences. There is a considerable amount of literature on sequence alignment tools and their advancements. The advancements reported in the literature relate to algorithms that seek to reduce running time and produce optimal alignments. Pairwise sequence alignment is best accomplished with the Dynamic Programming algorithm, which is slow and time consuming. Several 'shortcuts' to this algorithm have been proposed to improve running time. Best-known variants are the Smith–Waterman for local alignments and the Needleman–Wunsch for global alignments where sequences are related over their full length (Smith and Waterman 1981; Needleman and Wunsch 1970). These algorithms are, however, too compute time intensive to use for database searches. Most sequence databases allow rapid search using BLAST, FASTA (Altschul et al. 1990; Pearson 1990), scanps, MPsrch (http://www.ebi.ac.uk/searches/blitz_input.html); Blast2, PHI-Blast or BLAT (Kent 2002). BLAST is the fastest sequence alignment algorithm, although it compromises some degree of sensitivity in favor of speed. FASTA is slower, but more sensitive.

A multiple sequence alignment (MSA) is an alignment of three or more protein, DNA or RNA sequences and the purpose of creating such an alignment is to highlight their similarity or differences, which might reflect the biological relationship between them. Generation of MSA is a very useful exercise and needs special care when being used in phylogenetic tree construction, for identification of profiles and structure prediction, or in degenerate primer design. Computing exact MSAs is computationally almost impossible, and in practice approximate algorithms (heuristics) are used to align multiple sequences, by maximizing their similarity. Many MSA algorithms are in use, including the popular matrix-based methods ClustalW (Thompson et al. 1994) and Muscle (Edgar 2004), and the consistency-based methods T-Coffee (Notredame et al. 2000) and PCMA (Pei et al. 2003). Consistency-based methods are evaluated superior to matrix-based methods of alignment though they require cpu time several times higher than the matrix methods (Notredame and Abergel 2003). With the availability of so many quality methods and the growing importance of MSA generation, the development of meta-methods that can seamlessly combine the output of several methods, and also incorporate structure information, was the next milestone (Pei and Grishin 2006). Emerging advances in this area include template-based alignment, an extension of consistency-based methods. Under this new model, the purpose of an MSA is not to squeeze a dataset and extract all the information it may contain, but rather to use the dataset as a starting point for exploring and retrieving all the related information contained

in public databases. This information is used to drive the MSA computation. Such a usage of sequence and related information is seen as a major step toward global biological data integration (Notredame 2007). Jalview(http://www.jalview.org/download.html), BioEdit(http://www.mbio.ncsu.edu/BioEdit/BioEdit.html) and Genedoc(http://www.genedoc.us/gdsrc.htm) are popular freeware to edit multiple sequence alignments.

Several web-based tools are now available to browse and analyze genome alignments. These include the comparative genome viewers SynBrowse (Pan et al. 2005), SYBIL (http://sybil.sourceforge.net) and VISTA (Frazer et al. 2004). The VISTA family of tools includes a browser and rVISTA that combines a transcription factor binding site database search (using Blast) with comparative sequence analysis along with PHYLO-VISTA for phylogeny. Sybil is a web-based software package for comparative genomics, developed by the Bioinformatics group at J. Craig Venter Institute (formerly TIGR). This package includes several tools and browsers for genome comparisons and ortholog detection. FISH (Fast Identification of Segmental Homologies) is another useful algorithm available to explore the extent and distribution of conserved synteny between two species (Calabrese et al 2003). The Lagan Toolkit is a set of alignment programs for comparative genomics (Brudno et al. 2003) while the Staden Package is a suite of tools for sequence assembly, analysis, and mutation detection (Staden et al. 1998). The Gbrowse is a very popular viewer for manipulating and displaying annotations on genomes and was developed as part of the GMOD or Generic model organism database project. The tool is easy to use, fast, allows cross species comparisons, customizable and is freely available (http://www.gmod.org). The Ensembl Genome Browser is a software system using which a large selection of annotated eukaryotic genomes can be browsed and compared (http://www.ensembl.org/). Other comparative genomics tools include VisGenome (Jakubowska et al. 2007) and the SGN comparative map viewer (Mueller et al. 2008). cMAP (http://www.gramene.org/cmap/) allows comparisons of genetic and physical, sequence and QTL (Quantitative Trait Loci) maps, while CMTV (http://www.ncgr.org/cmtv/) allows comparative viewing of genetic and QTL maps and their integration to generate consensus maps.

4.2.5 Bioinformatics Tools for Functional Genomics

Functional genomics came of age when a shift of emphasis occurred from genome mapping and sequencing to determining how genes work together to produce traits. Current structural genomic approaches (i.e., mapping) generally focus on traits controlled by one or only a few genes, and often they provide information regarding the location of one or more genes only. Where functional information is available the scientist is equipped to a large extent to create varieties with exact combinations of traits. Most of available functional genomics resources are in the model crops, but since the genes that code for scores of plant traits and processes

are quite similar across many species, this knowledge can be applied to genetic research on other crops. Functional genomics as it is being applied in the plant sciences includes functional annotation, gene expression, and elucidation of protein structure that can help link genome and proteome with phenotype, protein–protein interaction, intracellular localization and posttranslational regulation. Rapid improvements in innovations such as microarray and RNA interference technology, allow simple, low-cost, high-throughput screening of phenotypes, as opposed to looking at just a few specific "candidate genes." The predominant methods for sequence-based expression analysis are SAGE (Serial Analysis of Gene Expression) and MPSS (Massively Parallel Signature Sequencing) of which SAGE is more widely used, while for model crops MPSS resources are available (http://mpss.dbi.udel.edu/).

Functional annotation is the process of collecting information about and describing a gene's biological identity – its various aliases, molecular function, biological role(s), subcellular location and its expression domains within the plant. The association between sequence and functional phenotype can be predicted using homology search tools based on sequence alignment. Larger data sources like TAIR (The Arabidopsis Information Resource) use a combination of published literature, solicited contributions from the research community as well as computational analyses of the sequence as part of the functional annotation process (Swarbreck et al. 2007). Pattern recognition programs, tools to transfer annotation to GO terms, as well as available controlled vocabulary add value to the annotation. Software such as *GeneTools*, allows users to rapidly extract gene annotation data, to add "user defined" GO annotation to gene products and to perform hypothesis testing using *e*GOn (Beisvag et al. 2006). B2GO is a single tool for the functional annotation of sequence data that uses BLAST to find homologous sequences to fasta formatted input sequences. The program extracts GO terms to each obtained hit and assigns GO terms to the query sequence using an annotation rule. Annotation and functional analysis can be visualized in graph form (http://www.blast2go.de/). Whichever the tool of choice, the user should be aware that the annotation is only an approximation that must be further validated computationally and/or through wet lab experimentation.

Existing open source software generated by the bioinformatics community for fragment assembly and mapping are well known and widely used (Phrap (http://www.phrap.com/), cap3 (http://genome.cs.mtu.edu/cap/cap3.html), PCAP (http://seq.cs.iastate.edu/) and TGICL(http://compbio.dfci.harvard.edu/tgi/software/)), while feature prediction tools like Genscan for gene structure prediction have versions suitable for crops such as maize and *Arabidopsis*. The NetPlantGene web server (http://www.cbs.dtu.dk/services/NetPGene/) provides tools for the prediction of splice sites in *Arabidopsis* besides modelling and structure prediction tools**.** The Database for Annotation, Visualization and Integrated Discovery (DAVID, http://david.abcc.ncifcrf.gov/) provides a comprehensive set of functional annotation tools. AutoFACT is another fully automated and customizable annotation tool that assigns biologically informative functions to a sequence (Koski et al. 2005). Other functional genomics platforms are also

becoming available such as the Purdue Ionomics Information Management System (PiiMS) that provides integrated workflow control, data storage and analysis to facilitate high-throughput phenotypic data acquisition, along with integrated tools for data search, retrieval and visualization for hypothesis development. PiiMS is deployed as a web-enabled system, allowing for integration of distributed workflow processes and open access to raw data for analysis by numerous laboratories (Baxter et al. 2007). This platform is currently being used to integrate high throughput phenotypic data with functional genomics data in *Arabidopsis*. TraitMill is an automated plant evaluation platform allowing high throughput testing of the effect of plant-based transgenes on agronomically valuable traits. The platform offers high throughput function prediction, allows selection of candidate trait improvement genes among annotated genes and is currently being used for rice (Reuzeau et al. 2006). The Generation Challenge Program (GCP) with the CGIAR centers, Advanced Research Institutes and a number of National Agricultural Research and Education Systems is also developing a platform for functional genomics customizing the MAXD database for rice gene expression data along with data mining and analysis pipelines (Takeya et al. 2006).

4.2.6 Availability of High Performance Clusters and Grid

The problems of biological datasets have only grown in scale and complexity with high throughput technology. Single experiments may generate gigabytes of data and a single gene product may have several thousand interactions that create more functions than one can imagine. So there is a continual demand for increased computation speed from a computer system. High performance compute (HPC) systems have been available since the mid-1970s to users with large budgets. For the others with limited budgets and large computing needs, hardware parallelism can be achieved by connecting several independent computers. The idea being that *n* computers can provide up to *n* times the computational speed of a single computer. The popular beowulf clusters are created through networking a group of computers running linux. Continual improvements in execution speeds of single processors and their availability has made such clusters faster and cheaper to build. There are a number of approaches available to creating effective parallel computers with different levels of effectiveness for different kinds of problems. For programmes to show an increase in speed a substantial fraction of the computation needs to be executed in parallel. Software parallelism is the ability to find well-defined areas in a problem that can be broken down into self contained parts. The distributed processing of these parts speeds the programme up. Such parallel programmes are increasingly being used in the agricultural domain for data mining, comparative genomics, phylogenetics and population genetics analysis applications as well as in breeding simulation programmes. Parallel systems are also being used for fault tolerant applications such as hosting distributed databases (high availability clusters). While with high performance clusters one can deploy a solution with a fixed number

of nodes (processors) on dedicated hardware, Grid computing brings several clusters together with the flexibility of using standard non-heterogeneous hardware where nodes can be added on demand and is not limited to the local LAN (Local Area Network), meaning that they could be geographically distributed. Through the Generation Challenge Progam, an HPC grid is becoming available that connects HPCs from four geographically distributed member institutions (http://hpc.cip.cgiar.org/webeval/), hosting several analysis software. Projects like myGrid allow biologists to design and execute in silico experiments on their desktop/laptop accessing datasources and tools available through the grid using the Taverna workflow bench. MyGrid uses the Feta web services discovery engine that is very similar in function to Moby Central of BioMoby (mygrid.org.uk).

4.2.7 Bioinformatics and Molecular Marker Technology

4.2.7.1 In silico Marker Mining Tools

Growing sequence information in databases has seen a corresponding increase in bioinformatics tools available to mine this information usefully. In the crop sciences, sequence data are useful sources of molecular markers like SSRs (simple sequence repeats), SNPs (single nucleotide polymorphisms), annotated ESTs, anchor markers, TRAPs (target region amplification polymorphisms), CISPs (conserved intron spanning primers) and conserved ortholog sets. Table 4.2 gives a compilation of the more popular tools available to researchers for the purpose of mining sequence data for putative molecular markers. Bioinformatics methods also allow the identification of functional markers that are more relevant and superior to random markers because they are linked to functional motifs and trait locus alleles. They rely on comparative genomics and phylogeny and elucidate the nature of genes conserved. Tools are available for the design of degenerate oligonucleotides for PCR for gene isolation and subsequent development of gene markers (Rose et al. 2003). The markers mined can then be applied to genetic trait mapping (Morgante and Salamini 2003). One can use the annotated genome of any one species to transfer knowledge to another genome. The identification of genes and related markers through computational methods is currently employed as a component of the marker development process.

4.2.7.2 Data Acquisition Software

Rapid data generation through high throughput methods has also led to the development of several systems for the capture, storage and retrieval of this data. Some freely available information management systems have been developed for genotyping, such as software to manage TaqMan SNP genotyping data (Monnier et al. 2005), the GenoDB (Li et al. 2001), AGL-LIMS (Jayashree et al. 2006b), PacLIMS (Donofrio et al. 2005) and SNPP (Zhao et al. 2005) each with different levels of

Table 4.2 Bioinformatics tools and pipelines available for in silico marker mining from sequence data

Tool	Marker	URL	Programming language
AutoSNP	SNP	http://www.cerealsdb.uk.net/discover.htm	Perl
CISPrimerTool	CISP	http://www.icrisat.org/gt-bt/softwares_downloads.htm	Java
GeMprospector	Cross species marker candidates	http://cgi-www.daimi.au.dk/cgi-chili/GeMprospector/main	Python, CGI
MISA	SSR	http://pgrc.ipk-gatersleben.de/misa	Perl
Polybayes	SNP	http://genome.wustl.edu/tools/software/polybayes.cgi	Perl
SNPdetector	SNP	http://lpg.nci.nih.gov	C and Perl
SNPpipeline	SNP	http://www.icrisat.org/gt-bt/softwares_downloads.htm	Parallel programme with an MPI wrapper (C++ and Python)
SSRIT	SSR	http://www.gramene.org/db/searches/ssrtool	Perl
Tandem Repeat Finder	SSR	http://tandem.bu.edu/trf/trf.download.html	Perl
TROLL (Tandem repeats occurrence locator)	SSR	http://sourceforge.net/projects/finder	C++

dependencies and functionalities. While GenoDB is a data management system for microsatellite markers and linkage analysis with functionalities tuned to human genotyping projects running on Windows platform, AGL-LIMS is a genotyping workflow management system for high throughput crop genotyping, platform independent and web enabled. Such systems, while serving as electronic notebooks for lab personnel, also help provide a measure of the quality of data being generated in the laboratory, better traceability and centralization of data. The ability to track data and communicate quality information gives the marker laboratory the tools to improve methods and work practices.

4.2.7.3 Molecular Marker Data Repositories and Visualization Tools

PlantMarkers is a genetic marker database that contains a comprehensive pool of predicted molecular markers (Rudd et al. 2005). The database contains putative single nucleotide polymorphism (SNP); simple sequence repeat (SSR) and conserved orthologue set (COS) markers. The database is derived from a systematic approach to identify a broad range of putative markers by screening the available openSputnik unigene consensus sequences from over 50 plant species. Cereal marker repositories include Gramene (Liang et al. 2008) and MaizeGDB (Lawrence 2008) while legume

marker repositories exist at LIS(Legume Information System) (Gonzales et al. 2005). Besides there are other multi-species marker databases published online as a result of individual institutional efforts such as the CUGI plant SSR database (http://www.genome.clemson.edu/projects/ssr/), SSRDB (Jayashree et al. 2006a) and TOGsDB (http://intranet.icrisat.org/gt1/tog/homepage.htm). The high-throughput marker discovery protocol – Diversity Arrays Technology (DArT) is sequence-independent (Jaccoud et al. 2001; Wenzl et al. 2004). As it becomes more accessible, there will soon be highly populated DArT marker databases. Major marker repositories also provide tools for the visualization of maps and comparisons with linkage maps from related species. The cereal markers repository Gramene provides cMAP, and the LIS allows the use of both cMAP and CMTV (Fig. 4.2). CMAP is

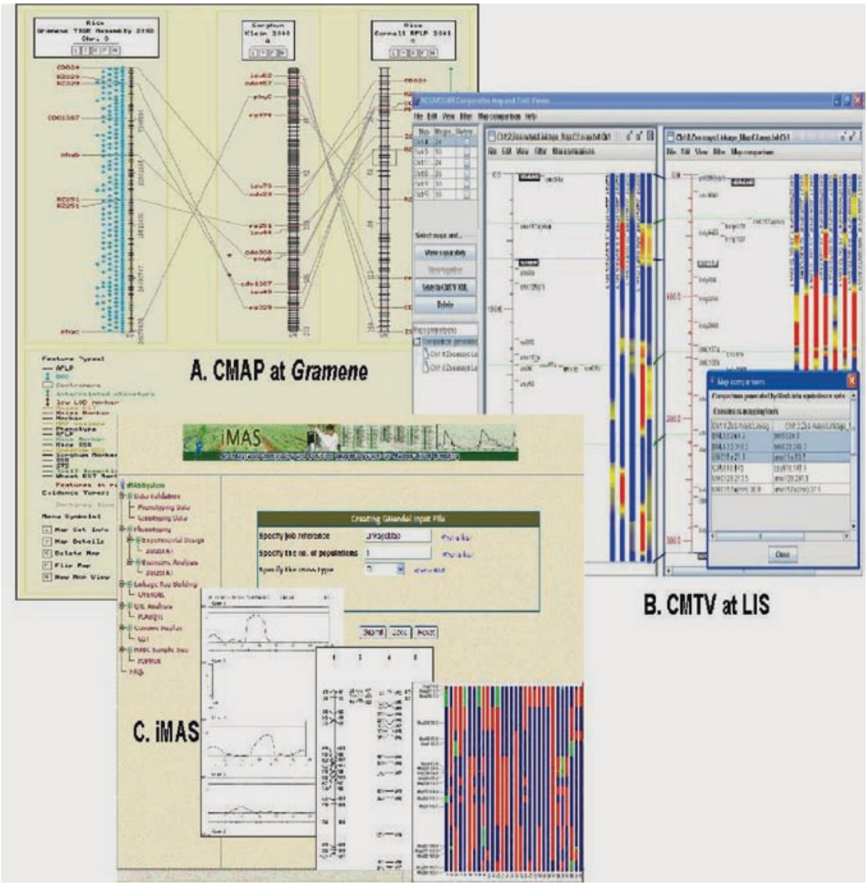


Fig. 4.2 Tools for map generation and comparisons. (**A**) The cMAP tool available at the gramene website (**B**) CMTV available from the Legume Information Service website (**C**) The desktop application iMAS

available under an open source license. CMTV allows viewing of multiple maps, the identification of correspondences as well as the combining of maps from different experiments to produce aggregate maps. This desktop application is also freely available (http://www.ncgr.org/cmtv/).

4.2.7.4 Software for Mapping and Association Analysis

The analysis of phenotypic and genotypic datasets leading to QTL maps, marker-aided selection and breeding involves the use of a number of different computing software. The last few years have seen a deluge of tools for map generation, association analysis and visualization. Many of these tools are available as freeware and some of them are open source (Table 4.3). There are several publications citing simulation software available to the plant breeder. Such tools have been used to investigate the introgression of one or several superior QTL alleles into a recipient line, to compare selection strategies based on proportion of recurrent parent genome recovered, and to investigate the effect of varying population size, marker density, marker positions, and required number of marker data points. Simulation approaches predict cross performance, compare different selection methods, and identify best performing crosses and breeding strategies. Software like PBMASS (pedigree-based marker assisted selection system) for MAS and recurrent parent recovery in wheat and barley has been published although the software is not publicly available (Eisemann et al. 2004).

4.3 Closing the Gap to Meet Molecular Breeding Requirements

Molecular breeding calls for integration of various kinds of information: genetic resource information with phenotype information linked to the allelic profiles of specific germplasm accessions coupled with results arising out of comparative and functional genomics experimentation. The goal is to rapidly assay the genetic makeup of individual plants or varieties in breeding populations and make accurate phenotypic predictions. This knowledge can be used to design a genoytpe that is targeted to perform well under a given set of environmental conditions. Marker assisted breeding programs typically involve information gathering over a prolonged period of time, need a management system to keep track of this information, and require a suite of analysis tools to help the scientist/breeder make decisions regarding which individuals to use from a segregating progeny. There is a need for systems that allow information to be carried forward and backward between the steps of the breeding program, allowing the user to choose breeding schema, to identify markers for foreground and background selection, to track inheritance and to serve as an information repository for data pertaining to the parental source materials, linkage maps, loci and genotyping data, polymorphism information for background and foreground markers in the parents and recombinants. Such systems will also serve as a link between the field books, the MAS and marker

Table 4.3 Software tools for mapping, association analysis and breeding simulation. The list is not extensive and includes only software available in the public domain

Tool	URL	Application
Adegenet	http://pbil.univ-lyon1.fr/software/adegenet	Related to ADE4, a R package for population genetics data analysis
Arlequin	http://cmpg.unibe.ch/software/arlequin3/	Implements a variety of population genetics methods that can be conveniently selected through the graphical interface
Blossoc	http://www.birc.dk/~mailund/Blossoc/	Linkage disequilibrium association mapping tool
CPSIM, BCSIM	http://www.plantbreeding.wur.nl/UK/software_cpsim.html	Simulation software for cross-pollinated population data or backcross simulation
GeneRecon	http://www.daimi.au.dk/~mailund/GeneRecon	LD mapping, based on a Bayesian MCMC method for fine scale linkage-disequilibrium gene mapping using high-density marker maps and association mapping
GGT	http://www.dpw.wau.nl/pv/PUB/ggt/	Graphical genotyping software
ICIM	http://www.isbreeding.net/software.html	Inclusive CIM, that provides an improvement over existing methods
IMAS	http://www.icrisat.org/gt-bt/download(bm)_iMAS.htm	Package of several integrated software for tasks from experimental design to map generation, qtl analysis and visualization along with a decision support platform
MADMAPPER	http://www.atgc.org/Xlinkage/MadMapper	Quality control of genetic markers, inference of linear order of markers on linkage groups
MAPL	http://lbm.ab.a.u-tokyo.ac.jp/software.html	QTL analysis by interval mapping and ANOVA, graphical genotyping
Mapmaker and Mapmaker/QTL	http://linkage.rockefeller.edu/soft/mapmaker/	QTL analysis, biologist friendly user interface
MapQTL	http://www.mapqtl.nl	Interval mapping, mapping QTLs for several types of mapping populations, Composite interval mapping, non-parametric mapping through a MS-Windows interface
PlabQTL	https://www.uni-hohenheim.de/plantbreeding/software/	Implement composite interval mapping besides others
PLABSIM	http://www.uni-hohenheim.de/~frisch/software.html	Plant breeding simulation software
PYPOP	http://www.pypop.org/	Software for the analysis of large-scale multi locus genotype data
QTLcartographer	http://statgen.ncsu.edu/	Implement composite interval mapping besides others
Qu-gene	http://www.uq.edu.au/lcafs/index.html?page=59974	Simulation platform for quantitative analysis of genetic models
Qu-Line	http://www.uq.edu.au/lcafs/index.html?page=59974	A component of Qu-gene, it is a simulation programme for the development of final advanced lines
STRAT	http://pritch.bsd.uchicago.edu/software/STRAT.html	Companion programme to Structure written for use in association mapping
Tassel	http://sourceforge.net/projects/tassel	Association mapping software

laboratory while providing easy to use interfaces and graphical visualization tools to view recombinant data. Thus, efficient use of DNA markers for crop improvement depends as much on computational tools as on laboratory technology. While information systems are becoming available for the acquisition, storage and retrieval of data derived from high throughput experimentation procedures, systems for integrating them with other data sources for the benefit of the plant breeder are as yet lacking. Software remains to be implemented that caters to the data integration needs of a plant breeder. Software specific to the management of information in marker assisted breeding programs is unavailable in the public domain. There is information available about the existence of LIMS for sample handling and databases specific to plant breeding operations, but these are private software packages developed for industry operated MAS programs that are neither licensed nor sold. Efforts are now being made to develop such information management systems (an ongoing project at ICRISAT).

For genomics to be applied to plant breeding, there is need for high throughput techniques, cost effective protocols, precise determination of quantitative trait expression, besides bioinformatics platforms that provide for the ability to combine outputs from these along with curated data on allelic variation annotated with alterations in phenotype. Thus, a high degree of curation for annotation polymorphisms with phenotypic variations in different genetic backgrounds is required along with high quality sequence annotation in selected germplasm resources. The Information Systems must also link to model crop data sources like genomic, genetic maps and functional genomics data sources. Figure 4.3 indicates the desired flow of information and integration of data sources. Crop improvement programmes can incorporate the results of genomics projects if they were available to those involved, namely the breeders. This calls for the coming together of a common platform for various disciplines at various locations. An example of one such succesfull disparate data/location integration initiative is PlaNet, a collaborative network of bioinformatics groups and plant molecular biologists from several plant genome data centres in Europe (JIC, NASC, CNB/CSIC, VIB, PRI and MIPS). The PlaNET approach to data integration reduces the strain on individual resources, distributes the burden of data curation and maximizes the value of individual data collections (Schoof et al. 2004). The established platform interconnects several databases, gathering external data into PlaNET through integration tools that allow flexible migration of data from various representations. This project uses BioMoby for interoperability. For crop improvement programmes to benefit from the various genomic resources and data collections, efforts such as these are needed that bring into the picture individual data sources held by groups that are involved in generating quality genotype, phenotype and genomics information for germplasm collections. Since curation is a long term effort, a consortium of dedicated data providers who are willing to share quality data across a common informatics platform accessible to breeders is a required investment. Careful annotation of DNA polymorphisms is required, whether the variation is indeed linked to an alteration in phenotype or whether it is a neutral sequence variation. The existing disparities in resources available to model crops research relative to orphan crops that are important to a large section of people in the developing world also needs to be closed. Increased investment in such crops will undoubtedly see a concomitant increase in bioinformatics

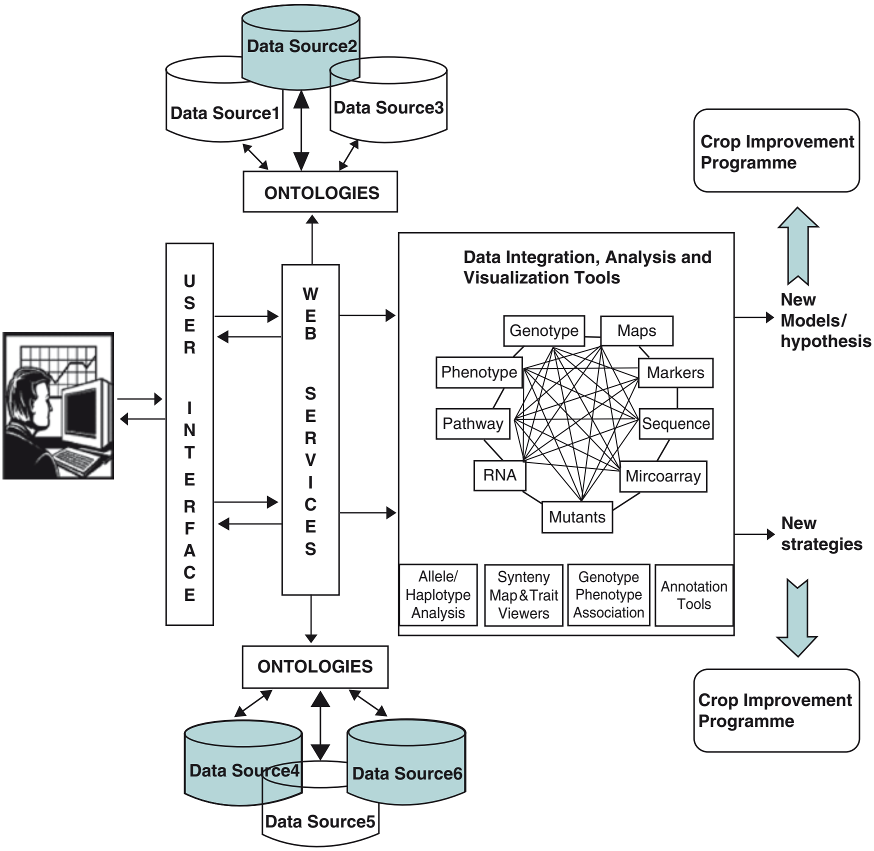


Fig. 4.3 Information and desirable data integration requirements for crop improvement programmes

data sources and adaptation/customization of tools developed for model crops. The availability of all this data through an integrated network of information to breeders who have been empowered to use it will provide the means to apply the outputs of modern technologies in crop improvement programmes.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman D.J (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Baxter I, Ouzzani M, Orcun S, Kennedy B, Jandhyala SS, Salt DE (2007) Purdue ionomics information management system. an integrated functional genomics platform. Plant Physiol 143:600–611

Beisvag V, Jünge FKR, Hallgeir B, Jølsum L, Lydersen S, Günther C-C, Ramampiaro H, Langaas M, Sandvik AK, Lægreid A (2006) GeneTools – application for functional annotation and statistical hypothesis testing. BMC Bioinform 7:470

Brudno M, Chuong Do, Cooper GM, Kim MF, Davydov E, Green ED, Sidow A, Batzoglou S (2003) LAGAN and Multi-LAGAN: Efficient tools for large-scale multiple alignment of genomic DNA. Genome Res 13:721–731

Calabrese PP, Chakravarty S, Vision TJ (2003) Fast identification and statistical evaluation of segmental homologies in comparative maps. Bioinformatics 19:i74–i80

Critchlow T, Fidelis K, Ganesh M, Musick R, Slezak T (2000) DataFoundry: information management for scientific data. IEEE Trans Inform Technol Biomed 4:52–57

Donofrio NM, Rajagopalan R, Brown DE, Diener SE, Windham DE, Nolin S, Floyd A, Mitchell TK, Galadima N, Tucker S, Orbach MJ, Patel G, Farman ML, Pampanwar V, Soderlund C, Lee Y-H, Deen RA (2005) PACLIMS: a component LIM system for high throughput functional genomic analysis. BMC Bioinformatics 6:94

Edgar RC (2004) Muscle: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

Eisemann B, Banks P, Butler D, Christopher M, Delacy I, Jordan D, Mace E, McGowan P, McIntyre L, Poulsen D, Rodgers D, Sheppard J (2004) Pedigree based genome mapping for marker assisted selection and recurrent parent recovery in wheat and barley. In: new directions for a diverse planet. Proceedings 4th International Crop Science Congress, Brisbane, 26 September–1 October, 2004)

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. Nucleic Acids Res 32:W273–W279

Gonzales MD, Archuleta E, Farmer A, Gajendran K, Grant D, Shoemaker R, Beavis WD, Waugh ME (2005) The Legume Information System (LIS): an integrated information resource for comparative legume biology. Nucleic Acids Res 33:D660–D665

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity Arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res 29:e25

Jakubowska J, Hunt E, Chalmers M, McBride M, Dominiczak AF (2007) VisGenome: visualization of single and comparative genome representations. Bioinformatics 23:2641–2642

Jayashree B, Crouch JH, Prasad PVNS, Hoisington D (2006a) A database of annotated tentative orthologs from crop abiotic stress transcripts. Bioinformation 1:225–227 (http://www.bioinformation.net/1/57–1–2006.htm)

Jayashree B, Reddy PT, Leeladevi Y, Crouch JH, Mahalakshmi V, Buhariwalla HK, Eshwar KE, Mace E, Folkertsma R, Senthilvel S, Varshney RK, Seetha K, Rajalakshmi R, Prasanth VP,Chandra S, Swarupa L, SriKalyani P, Hoisington DA (2006b) Laboratory Information Management Software for genotyping workflows: applications in high throughput crop genotyping. BMC Bioinformatics 7:383

Kent WJ (2002) BLAT – the Blast like alignment tool. Genome Res 12:656–664

Koski LB, Gray MW, Lang BF, Gertraud B (2005) AutoFACT: an automatic functional annotation and classification tool. BMC Bioinformatics 6:151

Lawrence CJ (2008) MaizeGDB, The maize genetics and genomics database. In: Methods in molecular biology: plant bioinformatics methods and protocols. Humana Press, Springer, pp 331–345

Lee TJ, Pouliot Y,Wagner V, Gupta P, Stringer-Calvert DWJ, Tenenbaum JD, Karp PD (2006) BioWarehouse: a bioinformatics database warehouse toolkit. BMC Bioinformatics 7:170

Li J-L, Deng H, Dong-Bing L, Fuhua X, Chen J, Gao G, Recker R, Deng H-W (2001) Toward high-throughput genotyping: dynamic and automatic software for manipulating large-scale genotype data using fluorescently labeled dinucleotide markers. Genome Res 11:1304–1314

Liang C, Jaiswal P, Hebbard C, Avraham S, Buckler ES, Casstevens T, Hurwitz B, McCouch S, Ni J, Pujar A, Ravenscroft D, Ren L, Spooner W, Tecle I, Thomason J, Tung CW, Wei X, Yap I, Youens-Clark K, Ware D, Stein L (2008) Gramene: a growing plant comparative genomics resource. Nucleic Acids Res (Database Issue) D947–D953

Monnier S, Cox DG, Albion T, Canzian F (2005) T.I.M.S: Taqman Information Management System, tools to organize data flow in a genotyping laboratory. BMC Bioinformatics 6:246

Morgante M, Salamini F (2003) From plant genomics to breeding practice. Curr Opin Biotechnol 14:214–219

Mueller L, Mills A, Skwarecki B, Buels R, Menda N, Tanksley S (2008) The SGN comparative map viewer. Bioinformatics Advance Access, published 17 January 2008

Needleman S, Wunsch C (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48:443–453
Nelson RJ, Naylor RL, Jahn MM (2004) The role of genomics research in improvement of "orphan" crops. Crop Sci 44:1901–1904
Notredame C (2007) Recent evolutions of multiple sequence alignment algorithms. PLoS Comp Biol 3:e123. doi:10.1371/journal.pcbi.0030123
Notredame C, Abergel C (2003) Using multiple alignment methods to assess the quality of genomic data analysis. In: Andrade M (ed) Bioinformatics and genomes: current perspectives. Horizon Scientific Press, Wymondham, UK, pp 30–55
Notredame C, Higgins DG, Heringa J (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol 302:205–217
Pan X, Stein L, Brendel V (2005) SynBrowse: a synteny browser for comparative sequence analysis. Bioinformatics 21:3461–3468. doi:10.1093/bioinformatics/bti555)
Pearson WR (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol 183:63–98
Pei J, Grishin NV (2006) MUMMALS: Multiple sequence alignment improved by using hidden Markov models with local structural information. Nucleic Acids Res 34:4364–4374
Pei J, Sadreyev R, Grishin NV (2003) PCMA: fast and accurate multiple sequence alignment based on profile consistency. Bioinformatics 19:427–428
Prlic A, Down TA, Kulesha E, Finn RD, Kahari A, Hubbard TJP (2007) Integrating sequence and structural biology with DAS. BMC Bioinformatics 8:333
Reuzeau C, Frankard V, Hatzfeld Y, Sanz A,Van Camp W, Lejeunne P, de Vilde C, Herve P, Peerbolte R, Broekaert W (2006) TraitMill: a functional genomics platform for the phenotypic analysis of cereals. Plant Genetic Resources: characterization and utilization 4:20–24
Rose T, Henikoff J, Henikoff S (2003) CODEHOP (COnsensus Degenerate Hybrid Oligonucleotide Primer) PCR primer design. Nucleic Acids Res 31:3763–3766
Rudd S, Schoof H, Mayer K (2005) PlantMarkers – a database of predicted molecular markers from plants. Nucleic Acids Res 33:D628–D632
Schoof H, Ernst R, Mayer KF (2004) The PlaNet consortium: A network of European plant databases connecting plant genome data in an integrated biological knowledge resource. Comp Funct Genomics 5:184–189
Smith AK, Kei-Hoi Cheung, Yip KY, Schultz M, Gerstein MB (2007) LinkHub: a Semantic Web system that facilitates cross-database queries and information retrieval in proteomics. Published in Semantic Web Approach to Database Integration in the Life Sciences. Baker CJO, Kei-Hoi Cheung (eds) Springer US, pp 11–30
Smith TF, Waterman MS (1981) Identification of common molecular subsequences. J Mol Biol 147:195–197
Staden R, Beal KF, Bonfield JK (1998) The Staden package. Meth Mol Biol 132:115–130
Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E (2007)The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Res. doi:10.1093/nar/gkm965
Takeya M, Suzuku K, Doi K, Kikuchi S, Bruskiewich R (2006) Development of a platform for functional genomics under the Generation Challenge Program. www.jsbi.org/modules/journal1/index.php/GIW06/GIW06P095.pdf
The BioMoby Consortium (2008) Interoperability with Moby 1.0 – It's better than sharing your toothbrush! Brief Bioinform 9:220–231
Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680
Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A, Kilian A (2004) Diversity Arrays Technology (DArT) for whole-genome profiling of barley. Proc Natl Acad Sci 101:9915–9920
Zhao L-J, Li M-X, Guo Y-F, Xu F-H, Li J-L, Deng H-W (2005) SNPP: automating large-scale SNP genotype data management. Bioinformatics 21:266–268

Part II
Molecular Markers and Their Application

Chapter 5
Gene-Based Marker Systems in Plants: High Throughput Approaches for Marker Discovery and Genotyping

Rajeev K Varshney

Abstract Development and application of molecular markers derived from genes, commonly called genic markers or sometimes functional markers, is gaining momentum in plant genetics and breeding. Availability of large amount of sequence data coming from genome/transcriptome sequencing projects as well as advent of next generation sequencing technologies together with advances in bioinformatics tools, marker discovery is becoming cheaper and faster. The availability of inexpensive high-density SNP-genotyping arrays is encouraging the plant genetics and breeding community to undertake genome-wide marker genotyping for a variety of applications. For instance, high-throughput and low cost genotyping assays for gene-based markers offers the possibility to accelerate the trait mapping based on high-density linkage mapping as well as genome-scanning based association mapping approaches in addition to facilitate physical mapping, comparative mapping, phylogenetic studies and understanding genome organization in crop plant species. Marker discovery, genotyping and molecular breeding practices would be routine in near future for crop improvement in many crop species. Advances in the area of marker discovery and genotyping using highly parallel genomics assays and also a few applications have been discussed in this chapter.

5.1 Introduction

Molecular markers are important genetic tools for plant breeders to detect the genetic variation available in the germplasm collection. During last two decades, varieties of molecular markers and in large numbers have been developed for almost all major crop species. Genetic variation detected by molecular markers has

R.K. Varshney
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru (AP), India; and
Generation Challenge Programme (GCP) CIMMYT, Mexico D.F, Mexico
e-mail: r.k.varshney@cgiar.org

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_5, © Springer Science+Business Media B.V. 2010

been useful for understanding the genome dynamics as well improving the breeding efficiency. For instance, these markers have been utilized extensively for the preparation of saturated molecular maps (genetic and physical) and their association with genes/QTLs controlling the traits of economic importance has been utilized in several cases for marker assisted selection (MAS) (Varshney et al. 2005b, 2006). As a result of extensive efforts undertaken at international level to identify molecular markers tightly linked with a large number of agronomic traits as well as tolerance/resistance to abiotic and biotic stresses in major crop species, it has been possible to realize the potential of molecular markers to track loci and genome regions in several crop-breeding programmes (Gupta and Varshney 2004; Varshney et al. 2006, 2007b). Other important uses of molecular markers include germplasm characterization, genetic diagnostics, genome organization studies and phylogenetic analysis (see Jain et al. 2002; Varshney and Tuberosa 2007).

Classically, the molecular markers can be grouped in three main categories (Gupta et al. 2002): (1) hybridization-based markers: restriction fragment length polymorphism (RFLP), (2) PCR-based markers: random amplification of polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP) and microsatellite or simple sequence repeat (SSR), and (3) sequence or chip-based markers: single nucleotide polymorphism (SNP), diversity array technology (DArTs) and single feature polymorphism (SFP). Indeed, till recent past, molecular markers from the first two categories have been developed in majority of the crop species, which belonged to genomic DNA, and therefore could belong to either the transcribed or the non-transcribed part of the genome without any information available on their functions. Nevertheless, during last few years, it has been possible to develop the markers from transcribed region of the genome or genes. Such markers have been popularly referred as functional markers/FMs (Anderson and Lübberstedt 2003; Gupta and Rustgi 2004), genic molecular markers/ GMMs (Varshney et al. 2007c) and gene expression markers/ GEMs (West et al. 2006b). Although development of gene-based markers is currently restricted to only limited crop species, the next generation sequencing technologies available very recently are enabling development of gene-based markers even in "orphan" crop species that are deficient in genomic resources (see Varshney et al. 2009). The present chapter deals with the advances made recently in the area of development of gene-based markers and methods of genotyping in crop species.

5.2 Gene-Based Marker System: Moving from Genes to Genome

As a result of establishment of several large scale genome/transcriptome sequencing and gene discovery projects in several plant species, a large number of genes have been identified through *wet lab* as well as *in silico* studies and a wealth of sequence data have been accumulated in public databases (e.g. http://www.ncbi.nlm.nih.gov; http://www.ebi.ac.uk) in the form of BACs (bacterial artificial chromosomes), ESTs (expressed sequence tags), GSSs (genome survey sequences), full length cDNA clones and genes. Furthermore, excellent progress has been made in

the area of development of bioinformatics tools and databases (see the companion chapter in this volume by Jayashree and Hoisington). Because of these two main factors, the era of development of gene-based markers has taken off in plant systems and is in advanced stage at present (Gupta and Rustgi 2004; Varshney et al. 2005a, 2007c, 2007d).

Based on the origin (polymorphic or non-polymorphic site in the gene), genic markers have been classified into two groups (Anderson and Lübberstedt 2003): (a) gene targeted markers (GTMs), that are derived from polymorphisms within genes, however not necessarily involved in phenotypic trait variation, e.g. EST-based molecular markers (Schmitt et al. 2006); (b) functional markers (FMs) are derived from polymorphic sites within genes causally involved in phenotypic trait variation, e.g. candidate gene-based molecular markers. Depending on the involvement in the phenotypic trait variation, the FMs can be classified further into two subgroups: (a) direct functional markers (DFMs), for which the role for the phenotypic trait variation is well proven, and (b) indirect functional markers (IFMs), for which the role for phenotypic trait variation is indirectly known (Anderson and Lübberstedt 2003).

5.3 Marker Discovery

Several types of molecular markers can be developed from genes and therefore can be grouped under genic marker category. Since several review articles (e.g. Gupta and Rustgi 2004; Varshney et al. 2005a, 2007c) have discussed at length the development of cDNA or EST-based RFLP, SSR and SNP markers, this article would mainly discuss the development of genic markers from sequence and expression data.

5.3.1 Sanger Sequencing-Based Marker Development

As a result of sequencing the genomes and transcriptomes followed by annotation provided the entire/partial gene repository of several model plant species (e.g. *Arabidopsis*, *Medicago*, *Lotus*, *Poplar*) and major crop species (rice, wheat, maize, sorghum, soybean, grape). In such plant species, based on functional annotation, gene expression and physiological studies, candidate genes can be identified for marker development. However in several crop species e.g. rye, sweet potato, pigeonpea and millets that have less genomic resources, identification of candidate genes is not easily feasible. However, because of the availability of next generation sequencing technologies, it has been shown very recently that the genic markers can be developed in so-called orphan crop species also (Varshney et al. 2009).

5.3.1.1 Candidate Gene-Based Marker Discovery

Based on positional cloning and/or knowledge of plant metabolic cycles, information on candidate gene(s) involved in the expression of a particular phenotype, is

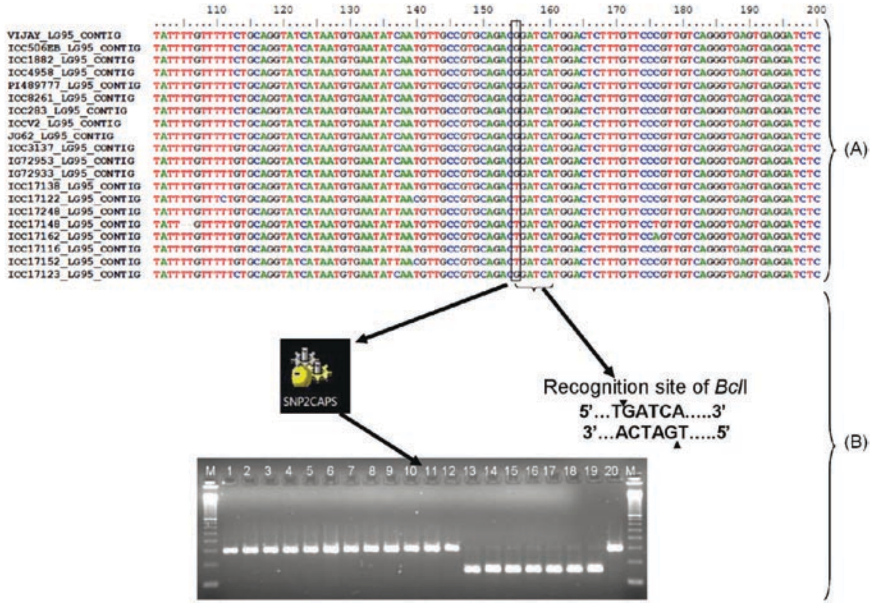


Fig. 5.1 An example of SNP discovery and conversion of SNP into CAPS assay. Multiple sequence alignment for sequence data generated for 20 genotypes using LG95 marker has been shown in panel A. This alignment reveals 7 SNPs at positions 111 (T/C), 113 (C/G), 137 (C/T), 140 (T/C), 155 (G/T), 176 (C/T/A) and 179 (T/C) in a window of 100 bp. SNP2CAPS analysis of these sequence data using SNP Pipeline developed at ICRISAT predicted the recognition site for restriction enzyme *Bcl*I at position 155. To verify the in silico identified restriction site involving SNP at 155 position, the PCR products of 20 genotypes generated by using LG95 marker were digested with restriction enzyme *Bcl*I. As a result, the PCR products, containing T nucleotide at 155 position in 7 genotypes were digested, while the amplicons, containing G nucleotide at the same position in remaining 13 genotypes remained intact (panel B). Thus the SNP at 155 position could be easily genotyped on agarose gel using CAPS assay (unpublished results).

available in many cases in model or major crop species. The primer pairs designed for the candidate gene(s) can be used to generate the amplicon in several genotypes of a species. Sanger dideoxy-nucleotide sequencing of these amplicons provides the allele specific sequencing data which can be subjected for multiple sequence alignment (MSA) by using bioinformatic tools and analyzed for identification of SNPs in the candidate gene (Fig. 5.1).

5.3.1.2 EST-Based Marker Discovery

For majority of main crop species, a large number of ESTs generated from several genotypes of a species are available in the sequence databases. The redundant set of ESTs for a given species can be used for *in silico* SNP discovery by employing bioinformatic scripts and tools for SNP discovery. A large number of bioinformatics tools or pipelines are available in public domain for identification of SNPs

Table 5.1 Some important SNP related tools and databases

Tool name	Web address	Description	Reference
SNP discovery tools			
SNP-VISTA	http://genome.lbl.gov/vista/snpvista/	A programme for visualization of mutations/SNPs in genes and discovery of recombination points	Shah et al. (2005)
SNPselector	http://primer.duhs.duke.edu/	A web tool for selecting SNPs for genetic association studies and it works on data available from Ensembl	Xu et al. (2005)
Forage	http://biobase.biotech.kth.se/forage/	A software application for SNP discovery which is based on two neural networks that have been trained on a set of validated SNPs	Unneberg et al. (2005)
SNPHunter	http://www.hsph.harvard.edu/ppg/software.htm	A software programme that allows for both ad hoc-mode and batch-mode SNP search, automatic SNP filtering, and retrieval of SNP data, including physical position, function class, flanking sequences at user-defined lengths, and heterozygosity from NCBI dbSNP	Wang et al. (2005)
SNPServer	http://hornbill.cspp.latrobe.edu.au/snpdiscovery.html	A real-time flexible tool for the discovery of SNPs and insertions/deletions	Savage et al. (2005)
SEAN	http//zebrafish.doc.ic.ac.uk/Sean	An application that predicts SNPs using multiple sequence alignments produced from EST clusters; the algorithm uses rules of sequence identity and SNP abundance to determine the quality of the prediction	Huntley et al. (2006)
QualitySNP	http://www.bioinformatics.nl/tools/snpweb/	A pipeline for detecting single nucleotide polymorphisms and insertions/deletions in EST data from diploid and polyploid species	Tang et al. (2006)
MAVIANT (Multipurpose Alignment VIewing and Annotation Tool)	http://snp.agrsci.dk/maviant/	A platform-independent tool that provides DNA chromatogram and alignment views and facilitates evaluation of predictions	Panitz et al. (2007)

(continued)

Table 5.1 (continued)

Tool name	Web address	Description	Reference
MSQT (Multiple SNP Query Tool)	http://msqt.weigelworld.org/	MSQT extracts SNP information from multiple sequence alignments, stores it in a database, provides a web interface to query the database and outputs SNP information in a format directly applicable for SNP-assay design	Warthmann et al. (2007)
QuickSNP	http://bioinformoodics.jhmi.edu/quickSNP.pl	An automated web server for selection of tagSNPs	Grover et al. (2007)
WASP	http://bioinfo.biotec.or.th/WASP	A tool for designing allele specific primers for both SNPs and mutations. By integrating the database for known SNPs, this tool facilitates the awkward process of getting flanking sequences and other related information from public SNP databases	Wangkumhang et al. (2007)
Haplo*SNP*er	http://www.bioinformatics.nl/tools/haplosnper/	A flexible web-based tool for detecting SNPs and alleles in user-specified input sequences from both diploid and polyploid species	Tang et al. (2008)
SNP-PHAGE	http://bfgl.anri.barc.usda.gov/ML/snp-phage/	A SNP discovery pipeline with additional features for identification of common haplotypes within a sequence tagged site (Haplotype Analysis) and GenBank (-dbSNP) submissions	Matukumalli et al. (2006)
Tools to convert SNPs to CAPS or dCAPS			
SNP2CAPS	http://pgrc.ipk-gatersleben.de/snp2caps/	A simple algorithm that involves the screening of multiply-aligned sequences for restriction sites followed by a selection pipeline that allows the deduction of CAPS candidates by the identification of putative alternative restriction sites	Thiel et al. (2004)

SNP Cutter	http://bioinfo.bsd.uchicago.edu/SNP_cutter.htm	The program is capable of designing primers for either natural PCR-RFLP or mismatch PCR-RFLP, depending on the SNP sequence data; SNP Cutter generates the information needed to evaluate and perform genotyping experiments, including a PCR primers list, sizes of original amplicons and different allelic fragment after enzyme digestion	Zhang et al. (2005)
SNP pipeline	http://www.icrisat.org/gt-bt/download_SNP.htm	A pipeline of open source to mine large EST datasets for SNPs and identify restriction sites for assaying the SNPs so that cost-effective CAPS assays can be developed for SNP genotyping.	Jayashree et al. (2007)
dCAPS Finder	http://helix.wustl.edu/dcaps/dcaps.html	The tool generates mismatches in a PCR primer that are used to create a polymorphism based on the target SNP and the tool identifies the restriction enzymes to genotype the SNP using CAPS assay	Neff et al. (2002)

(Table 5.1). Basically, all these tools/pipelines perform clustering on redundant set of ESTs and the ESTs representing one gene are grouped under one cluster, which can be visualized for the occurrence of SNPs. The SNP(s) identified by using this approach, however, should be verified either by checking the sequence chromatogram of ESTs or validated by *wet lab* experiments, as *in silico* SNPs may have been observed as a result of sequencing error.

5.3.1.3 Intron Targeted Marker Development by Using Comparative Genomics Approach

Sequence diversity projects undertaken in several crop species using large number of genes (and different parts of gene) suggested higher frequency of SNP in intronic regions of the gene (Ching et al. 2002; Rajesh and Muehlbauer 2008). Therefore the markers developed from intronic region in general show high level of polymorphism (Bertin et al. 2005). In case of well-characterized species, exon-intron boundaries can be defined using tools like FEGNESH, etc. and subsequently primer pairs can be designed using the flanking exonic sequence of an intron to amplify intronic region (Feltus et al. 2006).

By using the genome sequence data of *Medicago* and *Lotus* and EST data of soybean and *Medicago*, >3,000 primer pairs targeting intronic regions have been developed in the laboratory of Doug Cook at UC-Davis, USA (DR Cook, personal communication) to develop gene-based conserved orthologous sequence (COS) markers in legume species. These primer pairs have been used to amplify the intronic regions in the parental genotypes of mapping populations of seven legume species including chickpea, common bean, groundnut, cowpea and pigeonpea which are being sequenced at present to identify the SNPs between parental genotypes of the mapping populations of these species. This study is expected to develop the extended version of comparative legume genetic maps developed earlier by Choi et al. (2004). Development of intron targeted markers has been successful even in under-resourced crop species. In such cases, the ESTs of the targeted species are BLASTed against the genome sequence of closest model genome sequence data to identify the exon–intron boundaries. In this way, intronic sequence can be identified and exonic sequence of the ESTs are used to design the primer pairs to amplify the intronic region. Based on this concept, after aligning the ESTs of sorghum, pearl millet, *Allium* and *Musa* with rice genome, >3,600 primer pairs, called conserved intron spanning primers (CISPs) have been developed for monocot species (Feltus et al. 2006; Lohithaswa et al. 2007; http://www.plantgenome.uga.edu/CISP/). Following the similar approach, a larger number of gene-based markers have been developed and used for diversity and mapping studies in pearl millet (Bertin et al. 2005), lupin (Phan et al. 2007), etc. Recently developed “cisprimerTOOL” at ICRISAT (http://www.icrisat.org/gt-bt/CISPTool.htm) for the identification of conserved intron scanning regions using EST alignments to a completely sequenced model crop genome and designing conserved intron scanning primers will greatly facilitate development of CISPs in several orphan crop species (Jayashree et al. 2008).

5.3.2 Expression Polymorphism-Based Markers

Comprehensive gene expression platforms are available in several crop species that has made it possible to undertake transcriptome profile of different tissues of the same genotype or same tissue of different genotypes (Sreenivasulu et al. 2004; Krist and Yu 2007). Several recent studies have demonstrated the use of transcript abundance data from genomic DNA/reduced complexity genomic DNA or cRNA hybridizations to microarrays (Affymetrix) to reveal genetic polymorphisms, if the transcript profiling has been done on different genotypes of the species. This polymorphism, also called expression level polymorphism (ELP), has been used as marker to genotype individuals in mapping populations (West et al. 2006a, 2006b).

Initially, Affymetrix GeneChips were used for identifying ELPs (Winzeler et al. 1998). Affymetrix GeneChips basically contain 11 different 25 bp-oligos covering features of the transcribed regions of each of several thousand genes. Each of these features for every gene on the GeneChip is present as a so-called perfect match (PM) and mismatch (MM) oligonucleotide. While the PM exactly matches the sequence of a particular standard genotype (e.g. one parent of a mapping population), the MM differs from this in a single substitution in the central 13th base. Therefore, if the parental genotypes of a mapping population, used for expression study, differ in the amount of mRNA produced by the particular tissue, this should result in a relatively uniform difference in their hybridizations across the 11 features. Furthermore, if the parental genotypes produce the same amount of mRNA but contain a genetic polymorphism within their DNA which coincides with one particular feature (or overlapping features), this would also result in differential hybridizations, however confined to that feature alone. Such polymorphism observed has been termed as single feature polymorphism (SFP) (Borevitz et al. 2003). Majority of studies dealing with discovery and genotyping of SFP have been conducted in sequenced and well-characterized model species such as yeast, mouse and *Arabidopsis* (Brem et al. 2002; Borevitz et al. 2003; Bing and Hoeschele 2005; Bystrykh et al. 2005; Ronald et al. 2005; Kumar et al. 2007). Nevertheless, recently in large and complex genome species such as barley, Cui et al. (2005) and Rostoks et al. (2005) hybridized barley expression microarrays with cRNA, to reduce target complexity, and detected thousands of SFPs. To establish the sensitivity and specificity of SFP prediction in species with complex and unsequenced genomes, Luo et al. (2007) explored four methods for identifying SFPs from microarray experiments involving two tissues in two barley genotypes and their doubled haploid (DH) progeny. They identified >4,000 separate SFPs that accurately predicted the SNP genotypes of >98% of DH lines. Very recently, the use of cross species platform has been demonstrated for identification of SFPs. For instance, Das et al. (2008) hybridized soybean microarray with cRNA of cowpea and identified >1,000 SFPs in parental genotypes of mapping populations of cowpea.

5.3.3 Next Generation Sequencing Technologies for Genome-Wide Marker Discovery

Several crop genomes have already been sequenced or in advanced stage of sequencing, that has enhanced our understanding of genome architecture. However, such data will have limited relevance to many other important species which are generally distantly related to model organisms. Genome sequencing in non-model organisms has the potential to also be greatly enhanced by developments in sequencing technologies. Recent developments, due to growing interest in human genome re-sequencing, nucleic acid chemistry, nanotechnology and microscopy have led to a new generation of sequencing and genotyping technologies. These new technologies sequence DNA very fast and cheap, however in short fragments. These new methods are currently driving down sequencing costs and increasing capacity at an unprecedented rate that makes the whole genome resequencing possible (Hudson 2008; Mardis 2008; Gupta 2008).

At present, three main sequencing methods of next generation sequencing technologies are commercially available: (a) 454/FLX sequencing, (b) Solexa/Illumina 1 GB SBS (sequencing-by-synthesis) technology, and (c) AB SOLiD (Sequencing by Oligonucleotide Ligation and Detection) technology (Varshney et al. 2009). For all these DNA sequencing methods, genomic DNA is randomly sheared and individual DNA molecules are then immobilized on a solid support, which can be a microscopic bead (in this case one molecule is affixed to each bead) or a macroscopic support such as a flow cell or slide (in this case many molecules are arrayed randomly on the support) (Fan et al. 2006). Subsequently these individual DNA molecules are then amplified using the PCR. In case of bead-based methods, amplification is done in an emulsion phase where the beads are protected from cross-contamination by the barrier of an immiscible solvent. The polymerase colonies, often called 'polonies' or 'clusters', which are clonally identical DNA molecules either attached to a single bead or attached to a localized region on a solid support. While in bead-based methods, the beads are then either themselves immobilized on a planar support, or placed in individual microscopic wells, in case of non-bead-based methods the polonies are generated *in situ*. After producing a planar array of polonies, the respective sequencing chemistry is applied directly to the molecules on the support. Instead of separating elongation products, the sequence is interrogated at every base, by the use of either fluorescence or chemiluminescence to directly detect the incorporation of a base-specific chemical probe.

In addition to above mentioned three technologies, one new method called Single Molecule Sequencing, also known as third generation technology has received a great deal of attention and has potential to further increase throughput. Infact, various single-molecule and other sequencing and resequencing methods are under development in academic laboratories, and at several companies such as Biotage, Helicos, Li-Cor, Microchip Biotechnologies, Nanofluidics, Nanogen, Network Biosystems and Visigen (Hudson 2008). These next generation sequencing

technologies can be used for genome-wide marker discovery in both model/major and under-resourced crop species. Several bioinformatics tools and pipelines have been developed recently for analyzing the next generation sequence data for SNP discovery (Table 5.2). Indeed the bioinformatics community across the world is actively engaged in improving the tools for analyzing the next generation sequence data with higher accuracy and efficiency (Hillier et al. 2008; Smith et al. 2008; Varshney et al. 2009).

5.3.3.1 Re-Sequencing in Well-Characterized Species

In the species, which have genome or EST sequence data, genotypes of interest e.g. parental genotypes of mapping populations can be subjected for next generation sequencing technologies for genome-wide marker discovery. Re-sequencing can be done on cDNA population as well as genomic DNA (reduced representation genome) of different genotypes. Sequencing data generated using the next generation sequencing technologies can be aligned with the reference genome (genome/ transcriptome assembly). In such a way, genome-wide variants can be identified between the genotypes or compared to the reference genotype. For instance, based on 454 sequencing the transcriptomes of shoot apical meristems from two maize inbred lines namely B73 (260,000 reads) and Mo17 (280,000 ESTs), >36,000 SNPs were detected within 9,980 unique B73 genomic anchor sequences (Maize Assembled Genomic Islands, called MAGIs). Stringent post-processing reduced this number to >7,000 putative SNPs; over 85% (94/110) of a sample of these putative SNPs were successfully validated by Sanger sequencing (Barbazuk et al. 2007). Similar kind of SNP discovery projects using 454 and/or Solexa sequencing are underway in several other crop species like soybean (Hyten et al. 2008), chickpea (May et al. 2008), pigeonpea (ongoing studies at ICRISAT and NRCPB, India).

In case of rice, where the genome sequence is available, International Rice Research Institute (IRRI), in collaboration with its partners, employed array-based resequencing technology using very high-density oligomer arrays for genome-wide SNP discovery in 20 diverse varieties at Perlegen Biosciences (Kenneth McNally, personal communication). Under this project, 100 Mb of the Nipponbare rice genome (IRGSP release 4) corresponding to the fraction with little or no repetitiveness was chosen for SNP discovery. Perlegen designed 25-mer oligos with single base offsets tiled across the 100 Mb fraction of the genome for both strands with the 13th base in full degeneracy with each target position in the reference sequence interrogated by eight oligos. Subsequently, independent long-range PCR amplicons were produced for target pools across the regions arrayed on a particular wafer or chip. These LR-PCR amplicons were pooled, labeled and hybridized to the wafers. Based on Perlegen's model-based algorithms, 259,721 non-redundant SNPs have been predicted among 20 varieties. The estimated SNP frequency (2.6 SNPs per kb) is comparable to the figures obtained from pairwise comparisons of *indica* and *japonica* rice varieties.

Table 5.2 Important bioinformatic tools for analyzing next generation sequence data for marker discovery and gene expression studies

Tool	Web address	Description	Reference
Atlas-SNP	http://code.google.com/p/atlas-snp/	A tool for SNP/indel discovery from genome re-sequencing using next-generation sequencing technologies	Wheeler et al. (2008)
Next*GENe*	http://www.softgenetics.com/NextGENe.html	A software to analyze the next generation sequence data for de novo assembly, SNP/indel detection and transcriptome analysis	SoftGenetics, USA
PanGEA	http://www.kofler.or.at/Bioinformatics/PanGEA/index.html	The programme allows to map ESTs/ sequence tags to genes or whole genomes using algorithm which are especially adapted to next-generation sequencing technologies and mapping results may be displayed as a gene-expression-profile or used for SNP identification	Kofler et al. unpublished
Alpheus	http://alpheus.ncgr.org/	A web-based cyberinfrastructure platform for pipelining, visualization and analysis of GigaBase-scale sequence data generated from Sanger, Roche-454, Illumnia-Solexa, ABI SOLiD methodologies. Alpheus provides data management services, an analysis pipeline, and internet-accessible software for variant discovery and isoform identification.	National Centre for Genome Resource, USA
SNPsniffer	http://bioinformatics.bc.edu/marthlab/Polymorphism_Discovery_in_Next-Generation_Sequence_Data	A SNP discovery tool specifically designed for 454 sequences (currently being developed)	Aaron Quinlan and Weichun Huang, Boston College, USA

5.3.3.2 *De novo* Sequencing of Under-Resourced Crop Species

Although next generation sequencing technologies are ideally meant for re-sequencing, *de novo* sequencing can also be undertaken with these sequencing technologies. Alignment of the smaller fragments without the availability of the reference genome, however, becomes quite tedious if not impossible. In such case more than one genotype can be used for generating the sequence data using 454/Solexa/AB SOLiD technologies. Alignment of these sequence data can be facilitated by: (1) genome or transcriptome sequence data of model/major crop species closely related with the species; (2) whole transcriptome or reduced representation genome sequence data of the species, generated using 454 sequence technology. Aligning of sequence data of more than two genotypes of the species by using one of the above approaches provides the confidence in aligning the short sequence and detecting the sequence variants. Although several bioinformatic tools and algorithms are currently available (Table 5.2), efforts are continuously underway at several places to improve the accuracy of alignment of next generation sequence data (Smith et al. 2008). A web-based cyber infrastructure platform, called Alpheus (http://alpheus.ncgr.org/), is very useful for pipelining, visualization and analysis of GigaBase-scale sequence data for identification of SNPs.

A preliminary study dealing with Solexa sequencing of drought challenged root trasncriptomes of two genotypes of chickpea, ICC 4958 and ICC 1882 carried out at ICRISAT in collaboration with National Centre for Genome Resources (NCGR), USA (Greg May and Andrew Farmer) and University of California, Davis, USA (Doug Cook), has demonstrated the utility of next generation sequencing technology for SNP discovery in a species without the reference genome (May et al. 2008). Half run of Solexa sequencing on the pooled RNA samples from ICC 4958 and ICC 1882 yielded 5.2×10^6 and 3.6×10^6 sequence reads respectively. In order to analyze the generated Solexa datasets, the following three set of sequence resources were used in Alpheus pipeline: (1) *Medicago truncatula* (Mt) IMGAG (International *Medicago* Genome Annotation Group) gene assembly representing 29.5 Mb sequence data, (2) *Cicer arietinum* transcript assembly (Ca TA) of JCVI (The James Craig Venter Institute) representing 681 kb sequence data and (3) *Cicer arietinum* (Ca) BAC-end sequence (Ca BES) data representing 16.4 Mb sequence data. Bioinformatic analysis revealed matches of Solexa tags with 5,886 genes in cases of ICC 4958 and 7,338 genes in ICC 1882, respectively. Although detailed analysis for SNP discovery is underway, the preliminary analysis suggested the occurrence of SNPs at least in 500 cases.

5.4 Genotyping Assays

After identifying the SNPs in genes, optimizing or developing the appropriate platform for SNP genotyping is another important task. At present more than 30 SNP genotyping assays are available and each of them is having some merits as

well as constraints (Gupta et al. 2001, 2008). A critical comparison of a selected SNP genotyping assays has been made recently by Bagge and Lübberstedt (2008). Instead of discussing different kinds of assays in this article, some important and/or most commonly used SNP genotyping assays are discussed. These assays can be selected and optimized based on costs available and intended objective.

5.4.1 Low-Throughput and Inexpensive Genotyping Assay

When only few SNP-based genic markers need to be genotyped or limited financial resources are available, inexpensive SNP genotyping assays can be used. Different kinds of inexpensive SNP genotyping assays are currently available, only two assays have been given here.

5.4.1.1 Cleaved Amplified Polymorphic Sequences (CAPS)

Under this approach, the sequence alignment for more than two genotypes that contained SNPs is subjected to identify the restriction sites for restriction enzymes. This procedure can be facilitated by using bioinformatics tools, available in public domain (Table 5.1) that use multiple sequence alignments for several genes/markers in a batch file. The principle of these programmes is identification of recognition site and their corresponding restriction enzyme if the SNP present in multiple sequence alignment creates some recognition site for a restriction enzyme. Subsequently, the gene sequence can be amplified in germplasm through PCR and amplicon can be digested with the restriction enzyme identified by the programme and visualized on agarose gel. By using such methodology more than 80 EST-based SNP markers were converted into cost-effective CAPS markers (Kota et al. 2007). The approach has been used for assaying SNPs in many crop species, e.g. chickpea (Varshney et al. 2007d; Rajesh and Muehlbauer 2008), rye (Varshney et al. 2007a), rice (Komuri and Nitta 2005). Development of SNP markers and their optimization into CAPS assay is underway at ICRISAT (Reddy et al. unpublished). An example of such a CAPS-based genic SNP marker has been shown in Fig. 5.1.

In case, the SNP present in the genotypes of interest does not provide the recognition site for a restriction enzyme for CAPS assay, a modified technique called dCAPS (derived cleaved amplified polymorphic sequences) assay can be developed by creating a mismatch in a PCR primer to create a polymorphism based on the target mutation (Neff et al. 1998, 2002). In case of chickpea, genic SNP markers, which could not be converted to CAPS markers, are being assayed as dCAPS markers at ICRISAT (Reddy et al. unpublished).

5.4.1.2 Single-Strand Conformation Polymorphism (SSCP)

In general polyacrylamide gel electrophoresis doesn't allow detection of polymorphism due to difference of one base pair length/type. Single stranded DNA confirmation polymorphism (SSCP) methodology, however, allows detection of polymorphism due to differences of one or more base pairs in the PCR products that is suitable for SNP genotyping. The methodology relies on the secondary structure being different for single strands derived from PCR products that differ by one or more nucleotides at an internal site. For assays using SSCP methodology, PCR products of different genotypes carrying the SNP site are denatured and electrophoretically separated in neutral acrylamide gel. Because of occurrence of SNPs in different genotypes, the gel will show the difference in the length of the resolving fragments. This methodology has been used in several species like *Picea* (Germano and Klein 1999), pearl millet (Bertin et al. 2005), cassava (Castelblanco and Fregene 2006).

5.4.2 High-Throughput Genotyping Assays

Due to the availability of highly parallel genomic assays at present, large scale SNP-based marker genotyping is possible in cost-effective and relatively less time (Gupta et al. 2008). Majority of time, such genotyping assays are available as services offered by companies or genotyping centres.

5.4.2.1 GoldenGate Assay

GoldenGate assay of Illumina Inc. is probably the most popular large scale genotyping assay at present (Fan et al. 2003). The methodology deals with hybridization of allele (SNP)-specific primers directly to genomic DNA immobilised on a solid support. In case of a perfect match the primer is extended and the extension product is ligated to a probe hybridised downstream the SNP position. Subsequently, the ligated product is amplified by PCR using universal primers that are complementary to a universal sequence in the 3′end of the ligation probes and 5′ends of the allele-specific primers, respectively. It is important to note that the ligation probe contains a SNP-specific Tag-sequence while the universal allele-specific primers carry an allele-specific fluorescent label in their 5′end. After performing PCR, the amplified products are captured on beads carrying complementary target sequences for the SNP-specific Tag of the ligation probe. The beads are kept in fiber-optic array bundle that has a compatible format with 96-well microtiter plates. As a result, the GoldenGate assay supports genotyping of 96-, 192-, 384-, 768- and 1536 custom selected SNPs in a single reaction over a 3-day period.

Among plant systems, the GoldenGate assay-based SNP genotyping was undertaken for the first time in barley where barley community, in consultation with Illumina Inc., developed the GoldenGate assay for 1,536 SNPs selected based on EST mining (Rostoks et al. 2006). Inspired by high-throughput and low cost genotyping, the barley community has developed a total of three pilot Illumina oligonucleotide pool assays (OPAs) each containing 1,536 SNPs, under the barley coordinated agriculture project (CAP) in USA. These three OPAs have been used to map three mapping populations and genotype germplasm sets from the United States and Europe. From these three pilot OPAs, over 3,000 high quality SNPs have been used to design two OPAs (3,072 SNPs) for genotyping. These two OPAs will be used for genotyping 960 breeding lines of barley (http://www.barleycap.org/; TJ Close, personal communication). In soybean also, a custom 384-SNP GoldenGate assay was designed using SNPs discovered through the re-sequencing of five diverse accessions (Hyten et al. 2008). Allelic data were successfully generated for 89% of SNP loci (342 of the 384) and finally a gene-based integrated map with 334 SNP loci was prepared.

5.4.2.2 Whole-Genome Genotyping Infinium Assay

Based on Array-CGH (comparative genomic hybridization), Illumina Inc. introduced a very high-density SNP genotyping technology to genomic profiling, termed SNP-CGH, that allows simultaneous measurement of both signal intensity variations and changes in allelic composition. The utility of SNP-CGH was demonstrated with two Infinium whole-genome genotyping BeadChips, assaying 109,000 and 317,000 SNP loci, to detect chromosomal aberrations in samples bearing constitutional aberrations as well tumor samples at sub-100 kb effective resolution in human system (Peiffer et al. 2006).

Under the Infinium assay, *firstly* whole-genome amplification step is used to increase the amount of DNA up to 1,000-fold. Subsequently, the DNA is fragmented and captured on a BeadArray by hybridization to immobilised SNP-specific primers that is followed by extension with hapten-labelled nucleotides. As a result, the primers hybridize adjacent to the SNPs and are extended with a single nucleotide corresponding to the SNP allele. Finally, the incorporated hapten-modified nucleotides are detected by adding fluorescently labelled antibodies in several steps to amplify the signals. Data analysis under Infinium assays is performed using scatter plots as for the GoldenGate assay.

Very recently, Illumina Inc. announced development of the Infinium HD Human1M-Duo (two samples/chip) and the Human610-Quad (four samples/chip) system, featuring up to 2.3 million single nucleotide polymorphisms (SNPs) per BeadChip (www.illumina.com). Both arrays on the Human1M-Duo BeadChip contain markers for more than one million diverse genetic variants, all of which can be used for both whole-genome genotyping and copy number variation (CNV) analysis. The four-sample format of the Human610-Quad BeadChip offers a significant increase in sample throughput and reduced handling in the lab, as it has

550,000 SNPs plus an additional 60,000 genetic markers per sample. Although Infinium assay have not been developed in plant systems so far, availability of sequence data, next generation sequencing technologies for high density SNP discovery in some plant species like rice, maize, soybean may encourage the plant science community to undertake developing and using the Infinium assay soon.

5.5 Applications of Gene-Based Markers in Crop Improvement

Gene-based or functional markers (FMs) can be used for all the applications where traditional markers have been or can be used. In addition, the use of FMs provide added value for a particular application aimed at crop improvement. For instance, on one hand the FMs have been proven the "perfect markers" for foreground selection in marker-assisted selection (MAS), the availability of (or possibility to develop) low cost and high throughput genotyping platforms for gene-based markers (e.g. GoldenGate assays/Illumina arrays) for many crops make these markers the most suitable markers for background selection in marker-assisted breeding (MAB). When these markers are used in the genetic diversity studies, they assay the functional genetic variation in the germplasm collection and therefore can be used for allele mining and association genetics studies. Due to their origin from conserved proportion of the genome, gene-based markers of a species can be used in related species for a variety of applications including enhancing the density of genetic maps (Varshney et al. 2007a) and understanding the genome relationships and evolution (Stein et al. 2007). The utility of gene-based markers has been illustrated in selected two areas in following sections.

5.5.1 Superiority of FMs over Traditional Markers in MAS

Since 1990s, molecular markers have shown their applications for MAS in several crops (see Jain et al.2002 ; Gupta and Varshney 2004; Varshney and Tuberosa 2007). Large scale deployment of molecular markers in public breeding programme was initiated in 1997 at the Australian Molecular Plant Breeding Cooperative Research Centre (MPB CRC) in case of wheat (Eagles et al. 2001; Langridge 2005; Varshney et al. 2007b), shortly afterwards, such programmes were started in USA (MASwheat, that has been transformed into Wheat CAP recently – http://maswheat.ucdavis.edu/), Europe and China. Majority of these programmes have been deploying the SSR or STS/SCAR (developed from RFLPs) markers closely linked to disease resistance genes and agronomic traits.

A large number of markers associated with QTLs/genes for resistance/tolerance to biotic/abiotic stress as well as agronomic traits are reported every year in the form of research publications. However transfer of markers to practical plant breeding,

in the same proportion has not taken place at all or took longer than expected (Tuvesson et al. 2007). As a result, lots of claims have been made about what great things genomics does, but very little has been put into application (Varshney and Tuberosa 2007). One reason for this is the reduced reliability of diagnostic value of linked markers due to genetic recombination between marker and target locus (Bagge and Lübberstedt 2008). Such recombination, majority of times, impairs transfer of marker information from experimental mapping population to unrelated breeding materials. However this is not the case with gene-based and especially FMs that are derived from polymorphic sites within gene coding sequences causally affecting phenotypic trait variation. As a result, the FMs, as compared to anonymous markers including SSRs that were considered as markers of choice till recently, are more reliable for identification and selection of favourable alleles, as absence of recombination between marker and target locus increases the diagnostic power of the marker in the marker-based selection of genotypes.

In view of above, the FMs have been considered as 'good translators' from genomic technologies into improved crop varieties (Thro et al. 2004; Bagge et al. 2007). For example, in case of wheat, a STS marker for polyphenol oxidase (PPO) activity developed from the EST of the *PPO* gene, was found to discriminate accurately between 233 Chinese varieties with low and high PPO activity (Sun et al. 2005). Gene specific markers for waxy starch were used to select the wheat materials in Australia that had the *wx-B* allele, which is associated with good Asian noodle quality (Murai et al. 1999).

As mentioned earlier, development of FMs requires functionally characterized genes, the identification of polymorphic/functional site that affect plant phenotype within the corresponding genes and the validation of association between DNA polymorphisms and trait phenotype (Bagge et al. 2007), FMs have been developed so far only for selected traits and in few crops. For example, in case of rice, the cloning of the gene *xa-5* underlying the bacterial blight resistance (Iyer and McCouch 2004) has made it possible to develop functional markers for *xa5*-mediated resistance (Iyer-Pascuzzi and McCouch 2007). Bagge et al. (2007) and Bagge and Lübberstedt (2008) have recently summarized the current status on cloning of genes in wheat and their potential for functional marker development. A list of cloned genes in some major cereals like rice, wheat, barley and sorghum is available in Varshney et al. (2006) that could be used to develop FMs. Recent advances in the area of genomics like next generation sequencing technologies and high-throughput genotyping platforms mentioned earlier should facilitate the development and application of FMs in several crop species in coming years (Varshney et al. 2009).

5.5.2 Utility of Gene-Based Markers for Allele Mining

Gene-based markers and especially FMs are a better resource for allele mining for the corresponding gene from which the markers developed. The main approaches for allele mining include TILLING (targeting induced local lesions in genomes, see

Till et al. 2007), EcoTILLING (see Till et al. 2007) and candidate gene/FM sequencing. While TILLING approach deals with identification of new alleles after screening the mutant population for the candidate gene/FM, the EcoTILLING and candidate gene/FM sequencing approaches identity the natural allelic variation in a germplasm collection. By using TILLING approach, Slade et al. identified 246 alleles each homoeologue in 1,920 allohexaploid and allotetraploid wheat individuals. These alleles encoded *waxy* enzymes ranging in activity from near wild type to null, and they represented more genetic diversity than had been described in the preceding 25 years. An example of use of EcoTILLING for allele mining can be seen in case of rice for a gene “putative ethylene-responsive element binding protein 3 (ERF3) that colocalized with a drought QTL as 136.6 cM on chromosome 1 (Wang 2005). EcoTILLING in 2 kb ERF3 regions showed 25 SNPs and 5 indels defining 9 haplotypes (hap1 to hap9). Furthermore a significant association was reported between hap9 and yield stability in *indica* rice group.

Sequencing of candidate gene for the corresponding FM in an appropriate germplasm collection is a comprehensive approach for allele mining (Varshney et al. 2005b). For majority of the cereal and legume crops, a vast collection of genomic resources is available in genebanks of international agricultural research centres (IARCs) and national genebanks in different countries. In order to utilize the germplasm of genebank in breeding programme in the efficient way, the manageable collection of germplasm called “core collection”, “mini-core collection”, “reference collection” etc. have been developed in several crop species (e.g., Varshney et al. 2007c). Sequencing of the corresponding gene(s) for the FM in these germplasm collections provides a range of alleles for the given FM/gene. Phenotyping of the germplasm collection for the respective trait and their analysis with allele data should provide the better alleles for the respective trait (Slade and Knauf 2005). Infact, such allele mining and association genetics approaches are being used in several crop species to link genetic diversity with trait phenotype (Ersoz et al. 2007). This will help breeders to move towards allele-based selection in their breeding programmes.

5.6 Conclusions and Prospects

While development of markers and genetic map was an expensive and time consuming task till few years ago, availability of gene/genome sequence data together with high-throughput marker discovery and genotyping platforms have made the development of genic markers easier and faster. For instance, availability of three OPA assays in barley has provided 2,801 mappable gene-based SNP markers. Indeed, because of the possibility of large scale genotyping (with 1,536 SNP markers) of European barley cultivars, it was demonstrated that whole genome scanning-based association mapping is feasible in self-pollinated species like barley (Rostoks et al. 2006). Similar kinds of mapping efforts were undertaken in

soybean, maize, wheat, etc. and underway in many other crop species. High density marker genotyping is proving useful on one hand to anchor genetic map and physical map, linkage-disequilibrium-based association mapping approach is becoming possible for trait mapping. Among gene-based markers, the FMs are the "perfect markers" for foreground selection in MAS and therefore development of genic markers and especially FMs in different crops in coming future will enhance the application of MAS in breeding programmes and also for allele mining for the corresponding gene(s) using the germplasm collection held in genebanks or mutant populations. Availability of high-throughput genome-wide and low cost genotyping platform provides opportunities to accelerate breeding practices through the use of markers in background selection during marker-assisted breeding.

Recent advances in nanotechnology, nucleic acid chemistry, computational biology and automation indicates that development and application of gene-based markers using high-throughput marker discovery and genotyping assay is still a relatively young field, and more exciting advances are expected in the future. One of the great promises of genic markers, using high-throughput approaches, is that the ability to carry out comprehensive genomic analyses easily, inexpensively, accurately and rapidly with high sensitivity should create a new generation of routine genomic tools to assist the crop breeding.

Acknowledgements Thanks are due to Generation Challenge Programme (GCP), National Fund of Indian Council of Agricultural Research, Pigeonpea Genomics Initiative under Indo-US Agricultural Knowledge Initiative (AKI) and Department of Biotechnology (Government of India) for funding the research of author on development of genic molecular markers in legume species at ICRISAT.

References

Anderson JR, Lübberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554–560

Bagge M, Lübberstedt T (2008) Functional makers in wheat: technical and economic aspects. Mol Breed 22:319–328

Bagge M, Xia X, Lübberstedt T (2007) Functional makers in wheat. Curr Opin Plant Biol 10:211–216

Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS (2007) SNP discovery via 454 transcriptome sequencing. Plant J 51:910–918

Bertin, Zhu JH, Gale MD (2005) SSCP-SNP in pearl millet-a new marker system for comparative genetics. Theor Appl Genet 110:1467–1472

Bing N, Hoeschele I (2005) Genetical genomics analysis of a yeast segregant population for transcription network inference. Genetics 170:533–542

Borevitz JO, Liang D, Plouffe D, Chang HS, Zhu T, Weigel D, Berry CC, Winzeler E, Chory J (2003) Large-scale identification of single-feature polymorphisms in complex genomes. Genome Res 13:513–523

Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. Science 296:752–755

Bystrykh L, Weersing E, Dontje B, Sutton S, Pletcher MT, Wiltshire T, Su AI, Vellenga E, Wang J, Manly KF, Lu L, Chesler EJ, Alberts R, Jansen RC, Williams RW, Cooke MP, de Haan G (2005) Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. *Nat Genet* 37:225–232

Castelblanco W, Fregene M (2006) SSCP-SNP-based conserved ortholog set (COS) markers for comparative genomics in cassava (*Manihot esculenta* Crantz). Genet Resour 24:229–236
Ching ADA, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski A (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3:19
Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, Cook DR (2004) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci 101:15289–15294
Cui X, Xu J, Asghar R, Condamine P, Svensson JT, Wanamarker S, Stein N, Roose M, Close TJ (2005) Detecting single-feature polymorphisms using oligonucleotide array and robustified projection pursuit. Bioinformatics 21:3852–3858
Das S, Bhat PR, Sudhakar C, Ehlers JD, Wanamaker S, Roberts PA, Cui X, Close TJ (2008) Detection and validation of single feature polymorphisms in cowpea (*Vigna unguiculata* L. Walp) using a soybean genome array. BMC Genomics 9:107
Eagles HA, Bariana HS, Ogbonnaya FC, Rebetzke GJ, Hollamby GJ, Henry RJ, Henschke PH, Carter M (2001) Implementation of markers in Australian wheat breeding. Aust J Agric Res 52(12):1349–1356
Ersoz ES, Yu J, Buckler ES (2007) Applications of linkage disequilibrium and association mapping in crop plants. In: Varshney RK, Tuberosa R (eds) Genomics-assissted crop improvement, vol I. Springer, The Netherlands, pp 97–119
Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen M, Steemers F, Butler SL, Deloukas P, Galver L, Hunt S, Mcbride C, Bibikova M, Rubano T, Chen J, Wickham E, Doucet D, Chang W, Campbell D, Zhang B, Kruglyak S, Bentley D, Haas J, Rigault P, Zhou L, Stuelpnagel J, Chee MS (2003) Highly parallel SNP genotyping. Cold Spring Harbor Symopsium Quantitaive Biology 68:69–78
Fan JB, Chee MS, Gunderson KL (2006) Highly parallel genomic assays. Nature 7:632–644
Feltus FA, Singh HP, Lohithaswa HC, Schulze SR, Silva TD, Paterson AH (2006) Acomparative genomics strategy for targeted discovery of single-nucleotide polymorphism and conserved-noncoading sequences in orphan crops. Plant Phys 140:1183–1191
Germano J, Klein AS (1999) Species-specific nuclear and chloroplast single nucleotide polymorphisms to distinguish *Picea glauca*, *P. mariana* and *P. rubens*. Theor Appl Genet 99:1–2
Grover D, Woodfield AS, Verma R, Zandi PP, Levinson DF, Potash JB (2007) QuickSNP: an automated web server for selection of tagSNPs. Nucleic Acids Res 35:W115–W120
Gupta PK (2008) Single-molecule DNA sequencing technologies for future genomics research. Trends Biotech 26:602–611
Gupta PK, Rustgi S (2004) Molecular markers from the transcribed/expressed region of the genome in higher plants. Funct Integr Genomics 4:139–162
Gupta PK, Varshney RK (2004) Cereal Genomics. Kluwer Academic, The Netherlands, pp 639
Gupta PK, Roy JK, Prasad M (2001) Single nucleotide polymorphism: a new paradigm for molecular markers technology and DNA polymorphism detection with emphasis on their use in plants. Curr Sci 80:524–534
Gupta PK, Varshney RK, Prasad M (2002) Molecular markers: principles and methodology. In: Jain SM, Ahloowalia BS, Brar DS (eds) Molecular techniques in crop improvement. Kluwer Academic, The Netherlands, pp 9–54
Gupta PK, Rustgi S, Mir RR (2008) Array-based high throughput DNA markers for crop improvement. Heredity 101:5–18
Hillier LW, Marth GT, Quinlan AR, Dooling D, Fewell G, Barnett D, Fox P, Glasscock JI, Hickenbotham M, Huang W, Magrini VJ, Richt RJ, Sander SN, Stewart DA, Stromberg M, Tsung EF, Wylie T, Schedl T, Wilson RK, Mardis ER (2008) Whole-genome sequencing and variant discovery in *C.elegans*. Nat Methods 5:183–188
Hudson M (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. Mol Ecol Resour 8:3–17
Huntley D, Baldo A, Johri S, Sergot M (2006) SEAN: SNP prediction and display program utilizing EST sequence clusters. Bioinformatics 22:495–496

Hyten DL, Song Q, Choi IY, Yoon MS, Specht JE, Matukumalli LK, Nelson RL, Shoemarker RC, Young ND, Cregan PB (2008) High-throughput genotypying with the GoldenGate assay in the complex genome of soybean. Theor Appl Genet 116:945–952

Iyer AS, McCouch SR (2004) The rice bacterial blight resistance gene *xa5* encodes a novel form of disease resistance. Mol Plant Microbe Interact 17(12):1348–1354

Iyer-Pascuzzi AS, McCouch SR (2007) Recessive resistance genes and the *Oryza sativa-Xanthomonas oryzae* pv. *oryzae pathosystem*. Mol Plant Microbe Interact 7:731–739

Jain SM, Brar DS, Ahloowalia BS (2002) Molecular techniques in crop improvement. Kluwer Academic, The Netherlands, pp 601

Jayashree B, Hanspal MS, Srinivasan R, Vigneshwaran R, Varshney RK, Spurthi N, Eshwar K, Ramesh N, Chandra S, Hoisington DA (2007) An integrated pipeline of open source software adapted for multi-CPU architectures: use in the large-scale identification of single nucleotide polymorphisms. Comp Funct Genomics 35604:7

Jayashree B, Jagdeesh VT, Hoisington D (2008) cisprimertool: software to implement a comparative genomics strategy for the development of conserved intron scanning (CIS) markers. Mol Ecol Resour 8:575–577

Krist M, Yu Q (2007) Genetical genomics: successes and prospects in plants. In: Varshney RK, Tuberosa R (eds) Genmics assisted crop improvement. Springer, The Netherlands, pp 245–266

Komuri T, Nitta N (2005) Utilization of the CAPS/dCAPS method to convert rice SNPs into PCR-based markers. Breed Sci 55:93–98

Kota R, Varshney RK, Prasad M, Zhang H, Stein N, Graner A (2007) EST-derived single nucleotide polymorphism markers for assembling genetic and physical maps of the barley genome. Funct Integr Genomics. doi:10.1007/s10142-077-0060-9

Kumar R, Qiu J, Joshi T, Valliyodan B, Xu D, Nguyen HT (2007) Single feature polymorphism discovery in Rice. PLoS ONE 2:e284

Langridge P (2005) Molecular breeding of wheat and barley. In: Tuberosa R, Phillips RL, Gale M (eds) In the wake of the double helix: from the green revolution to the gene revolution. Avenue Meida, Bologna, Italy, pp 279–286

Lohithaswa HC, Feltus FA, Singh HP, Bacon CD, Bailey CD, Paterson AH (2007) Leveraging the rice genome sequence for monocot comparative and translational genomics. Theor Appl Genet 115:237–243

Luo ZW, Potokina E, Druka A, Wise R, Waugh R, Kearsey MJ (2007) SFP Genotyping from affymetrix arrays is robust but largely detects *Cis*-acting expression regulators. Genetics 176:789–800

Mardis ER (2008) The impact of next-generation sequencing technology on genetics. Trends Genet 24:133–141

Matukumalli LK, Grefenstette JJ, Hyten DL, Choi IY, Cregan PB, Van Tassell CP (2006) SNP-PHAGE-High throughput SNP discovery pipeline. BMC Bioinformatics 7:468

May GD, Lekha PT, Kashiwagi J, Huntley JJ, Farmer AD, Cook DR, Varshney RK (2008) Whole transcriptome shotgun sequencing for variant detection and transcript profiling in chickpea (*Cicer arietinum* L.). In: Plant Animal Genomes XVI Conf San Diego USA P385 (http://www.intl-pag.org/16/abstracts/PAG16_P05f_385.html)

Murai J, Taira T, Ohta D (1999) Isolation and characterization of the three *Waxy* genes encoding the granule-bound starch synthase in hexoploid wheat. Gene 234:71–79

Neff MM, Neff JD, Chory J, Pepper AE (1998) dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. Plant J 14:387–392

Neff MM, Turk E, Kalishman M (2002) Web-based primer design for single nucleotide polymorphism analysis. Trends Genet 18:613–615

Panitz F, Stengaard H, Hornshøj H, Gorodkin J, Hedegaard J, Cirera S, Thomsen B, Madsen LB, Høj A, Vingborg RK, Zahn B, Wang X, Wang X, Wernersson R, Jørgensen CB, Scheibye-Knudsen K, Arvin T, Lumholdt S, Sawera M, Green T, Nielsen BJ, Havgaard JH, Brunak S, Fredholm M, Bendixen C (2007) SNP mining porcine ESTs with MAVIANT, a novel tool for SNP evaluation and annotation. Bioinformatics 23:i387–i391

Peiffer DA, Le JM, Steemers FJ, Chang W, Jenniges T, Garcia F, Haden K, Li J, Shaw CA, Belmont J, Cheung SW, Shen RM, Barker DL, Gunderson KL (2006) High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. Genome Res 16:1136–1148

Phan HT, Ellwood SR, Adhikari K, Nelson MN, Oliver RP (2007) The first genetic and comparative map of white lupin (*Lupinus albus* L.): identification of QTLs for anthracnose resistance and flowering time, and a locus for alkaloid content. DNA Res 14:59–70

Rajesh PN, Muehlbauer FJ (2008) Discovery and detection of single nucleotide polymorphism (SNP) in coding and genomic sequences in chickpea (*Cicer arietinum* L.) Euphytica. doi: 10.1007/s10681-008-9675-8

Ronald J, Akey JM, Whittle J, Smith EN, Yvert G, Kruglyak L (2005) Simultaneous genotyping, gene-expression measurement, and detection of allele-specific expression with oligonucleotide arrays. Genome Res 15:284–291

Rostoks N, Borevitz JO, Hedley PE, Russell J, Mudie S, Morris J, Cardle L, Marshall DF, Waugh R (2005) Single-feature polymorphism discovery in the barley transcriptome. Genome Biol 6 (6), art. no. R54

Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Graner A, Close TJ, Waugh R (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc Natl Acad Sci USA 103:18656–18661

Savage D, Batley J, Erwin T, Logan E, Love CG, Lim GAC, Mongin E, Barker G, Spangenberg GC, Edwards D (2005) SNPServer: a real-time SNP discovery tool. Nucleic Acids Res 33:W493–W495

Schmitt BA, Costa JH, de Melo DF (2006) AOX -A functional marker for efficient cell reprogramming under stress? Trends Plant Sci 11:281–287

Shah N, Teplitsky MV, Minovitsky S, Pennacchio LA, Hugenholtz P, Hamann B, Dubchak I (2005) SNP-VISTA: An interactive SNP visualization tool. BMC Bioinformatics 6:292

Slade AJ, Knauf VC (2005) TILLING moves beyond functional genomics into crop improvement. Transgenic Res 14:109–115

Smith AD, Xuan Z, Zhang MQ (2008) Using quality scores and longer reads improves accuracy of Solexa read mapping. BMC Bioinformatics 9:128

Stein N, Prasad M, Scholz U, Thiel T, Zhang H, Wolf M, Kota R, Varshney RK, Perovic D, Grosse I, Graner A (2007) A 1000 loci transcript map of the barley genome- new anchoring points for integrative grass genomics. Theor Appl Genet 114:823–839

Sreenivasulu N, Varshney RK, Kavikishore PV, Weschke W (2004) Tolerance to abioitic stress- a functional genomics approach. In: Gupta PK, Varshney RK (eds) Cereal genomics. Kluwer Academic, The Netherlands, pp 483–514

Sun DJ, He ZH, Xia XC, Zhang LP, Morris C, Appeis R, Ma W, Wang H (2005) A novel STS marker for polyphenol oxidase activities in bread wheat. Mol Breed 16:209–218

Tang J, Vosman B, Voorrips RE, van der Linden CG, Leunissen JA (2006) QualitySNP: a pipeline for detecting single nucleotide polymorphisms and insertions/deletions in EST data from diploid and polyploid species. BMC Bioinformatics 7:438

Tang J, Leunissen JA, Voorrips RE, van der Linden CG, Vosman B (2008) HaploSNPer: a web-based allele and SNP detection tool. BMC Genet 9:23

Thiel T, Kota R, Grosse I, Stein N, Graner A (2004) SNP2CAPS: a SNP and INDEL analysis tool for CAPS marker development. Nucleic Acids Res 32:e5

Thro AM, Parroot W, Udall JA, Beavis WD (2004) The experience of initiative for future agricultural and food systems. Crop Sci 44:1893

Till BJ, Comai L, Henikoff S (2007) Tilling and Ecotilling for crop improvement In: Varshney RK, Tuberosa R (eds) Genomics-assissted crop improvement, vol I, Springer, The Netherlands, pp 333–349

Tuvesson S, Dayteg C, Hagberg P, Manninen O, Tanhuanpää P, Tenhola-Roininen T, Kiviharju E, Weyen J, Förster J, Schondelmaier J, Lafferty J, Marn M, Fleck A (2007) Molecular markers and doubled haploids in European plant breeding programmes. Euphytica 158:305–312

Unneberg P, Strömberg M, Sterky F (2005) SNP discovery using advanced algorithms and neural networks. Bioinformatics 21:2528–2530
Varshney RK, Graner A, Sorrells ME (2005a) Genic microsatellite markers in plants: features and applications. Trends Biotech 23:48–55
Varshney RK, Graner A, Sorrells ME (2005b) Genomics-assisted breeding for crop improvement. Trends Plant Sci 10:621–630
Varshney RK, Hoisington DA, Tyagi (2006) Advances in cereal genomics and applications in crop breeding. Trends in Biotech 24:490–499
Varshney RK, Beier U, Khlestkina EK, Kota R, Korzun V, Graner A, Börner A (2007a) Single nucleotide polymorphisms in rye (*Secale cereale* L.): discovery, frequency, and applications for genome mapping and diversity studies. Theor Appl Genet 114:1105–1116
Varshney RK, Langridge P, Graner A (2007b) Application of genomics for molecular breeding of wheat and barley. Adv Genet 58:122–155
Varshney RK, Mahender T, Aggrawal RK, Börner A (2007c) Genic molecular markers in plants: development and applications. In: Varshney RK, Tuberosa R (eds) Genomics-Assisted Crop Improvement, Vol I: Genomics Approaches and Platforms. Springer, The Netherlands, pp 13–30
Varshney RK, Nayak S, Jayashree B, Eshwar K, Upadhyaya HD, Hoisington DA (2007d) Development of cost-effecive SNP assays for chickpea genome analysis and breeding. J SAT Agric 3:1. http://www.icrisat.org/Journal/chickpea_pigeonpea3.htm
Varshney RK, Tuberosa R (2007) Genomics-assisted crop improvement: an overview. In: Varshney RK, Tuberosa R (eds) Genomics-Assissted Crop Improvement, Vol I: Genomics Approaches and Platforms, Springer, The Netherlands, pp 1–12
Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next generation sequencing technologies and their implications for crop genetics and breeding. Trends in Biotech, in press
Wang HH (2005) Application of ecoTILLING to relate molecular variation in a rice ethylene response factor (ERF3) gene to drought stress responses. Los Baños, Laguna, Philippines.86 I
Wang L, Liu S, Niu T, Xu X (2005) SNPHunter: a bioinformatic software for single nucleotide polymorphism data acquisition and management. BMC Bioinformatics 6:60
Warthmann N, Fitz J, Weigel D (2007) MSQT for choosing SNP assays from multiple DNA alignments. Bioinformatics 23:2784–2787
Wangkumhang P, Chaichoompu K, Ngamphiw C, Ruangrit U, Chanprasert J, Assawamakin A, Tongsima S (2007) WASP: a Web-based Allele-Specific PCR assay designing tool for detecting SNPs and mutations. BMC Genomics 8:275
Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM (2008) The complete genome of an individual by massively parallel DNA sequencing. Nature 452:872–876
West MAL, Kim K, Kliebenstein, van Leeuwen H, Michelmore RW, Doerge RW, St Clair DA (2006a) Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in *Arabidopsis*. Genetics 175:1441–1450
West MAL, van Leeuwen H, Kozil A, Kliebenstein DJ, Doerge RW, St Clair DA, Michelmore RW (2006b) High-density haplotyping with microarray-based expression and single feature polymorphism markers in *Arabidopsis*. Genome Res 16:787–795
Winzeler EA, Richards DR, Conway AR, Goldstein AL, Kalman S, McCullough MJ, McCusker JH, Stevens DA, Wodicka L, Lockhart DJ, Davis RW (1998) Direct allelic variation scanning of the yeast genome. Science 281:1194–1197
Xu H, Gregory SG, Hauser ER, Stenger JE, Pericak-Vance MA, Vance JM, Züchner S, Hauser MA (2005) SNPselector: a web tool for selecting SNPs for genetic association studies. Bioinformatics 21:4181–4186
Zhang R, Zhu Z, Zhu H, Nguyen T, Yao F, Xia K, Liang D, Liu C (2005) SNP Cutter: a comprehensive tool for SNP PCR-RFLP assay design. Nucleic Acids Res 33:W489–492

Chapter 6
Automation of DNA Marker Analysis for Molecular Breeding in Crops

Christophe Dayteg and Stine Tuvesson

Abstract Plant breeders constantly need to adapt their research to the ever-changing market needs and agricultural practices. To achieve these goals, they need to competently combine different genetically-governed characters in a genotype, this is a complex, time-consuming and labour intensive task. In modern plant breeding, molecular markers are of increasing importance, and it is today undeniable that their application inhold tremendous possibilities to increase plant breeding efficiency. While the methods are more widely adopted, the capacity for high-throughput analyses at low cost becomes crucial for their practical use. To be attractive it is necessary that molecular technology is able to promptly handle sufficiently large amounts of material at reduced costs. Automation of the analysis processes is a way to meet these requirements. In that purpose, the specific needs of molecular applications in practical plant breeding are investigated in this chapter. The particular approach of a plant breeding company to automate them, in order to increase their availability to breeding programs, is described.

6.1 Introduction

Traditionally, crop improvement has been achieved through the selection of observable phenotypes, representing the collective effect of all genes and the environment. This is a time consuming effort that is largely dependant on the performance of the selected candidates under certain environmental conditions. It is limited by the necessity that the phenotype has to be observable before the time when selection decisions have to be made or by its effectiveness in resolving negative association between genes. Hence, plant breeders' great interest in technologies

C. Dayteg (✉)
Department of Plant Breeding and Biotechnology, Swedish University of Agricultural Sciences, Box 44, SE-23053, Alnarp, Sweden
e-mail: cdayteg@gmail.com

S. Tuvesson
Svalöf Weibull AB, SW Laboratory, SE-26 881, Svalöv, Sweden
e-mail: stine.tuvesson@swseed.com

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_6,

that could make this procedure more efficient (Dekkers and Hospital 2002; Korzun 2003). Molecular markers are now a well-established technology that can be used in large breeding programmes to complement the traditional breeding process (Tuvesson et al. 2007). This chapter describes the automation of molecular marker analysis at the marker laboratory of Svalöf Weibull AB (SW) and is based on results published in a Ph.D. thesis: "Automation of Molecular Marker in Practical Breeding of Spring Barley *(Hordeum vulgare* L.)" (Dayteg 2008).

6.2 Plant Breeding and Molecular Markers

The exploitation of factors co-segregating with a trait in a simple Mendelian fashion in order to understand its inheritance is an old notion, but these simply inherited morpho-physiological variants are very rare[1] (Bergal and Friedberg 1940). They remained of restricted use for practical breeding purposes until the development of biochemical markers in the 1960s (Koebner 2003). However, it was not until the introduction of DNA marker technology in the 1980s, that a large enough number of environmentally insensitive genetic markers could be generated. Restriction fragment length polymorphisms (RFLPs) were the first DNA markers to be successfully used in plants (Helentjaris et al. 1985). However, as these markers are time-consuming, labour-intensive and require large amounts of DNA, their use was gradually supplanted by more user-friendly techniques (Gupta et al. 1999). Indeed, the development of the polymerase chain reaction (PCR, Saiki et al. 1988) has made DNA marker-techniques quicker and cheaper. Several PCR-based markers such as random amplified polymorphic DNAs (RAPDs), amplified fragment length polymorphisms (AFLPs), simple sequence repeats (SSRs or microsatellites), inter-SSRs (ISSRs) and most lately single nucleotide polymorphism (SNP) have been developed and applied to a wide range of crop species including cereals. The relative pros and cons of these techniques are summarized in Table 6.1.

Under the past decades, the molecular marker technology has rapidly evolved into a valuable tool able to dramatically enhance the efficiency of conventional plant breeding (Peleman and van der Voort 2003). Its various uses, as exemplified in the different chapters of this edition, have given modern crop improvement new potentials unthinkable until recently i.e. association mapping, induced mutagenesis, Tilling etc. Simultaneously as biotechnology produces these efficient tools to assist plant breeders in their enterprise, it also provides them with new possibilities of gene transfer. To breed and/or distinguish genetically modified (GM) individuals, may not differ much from other traits, however, molecular markers is the only

[1]In 1875 von Proskowetz used ear selection as a predictor of malting quality. Tolerance to pre-harvest sprouting in wheat was known to be associated with red kernels (Nilsson-Ehle 1914) allowing indirect selection for this trait. In the 1920's simple colour traits were used to predict seed weight in common bean, and fruit size in tomato. Others were used for varietal discrimination (DUS).

Table 6.1 Comparison of the most common used marker systems in crops. Adapted from Korzun (2003)

Feature	RFLPs	RAPDs	AFLPs	ISSRs	SSRs	SNPs
PCR-based	No	Yes	Yes	Yes	Yes	Yes
Ease of use	Low	High	High	High	High	High
Number of polymorphic loci assayed	1–3	1–50	10–100	5–30	1–3	1
Reproducibility	High	Low	High	High	High	High
Amenable to automation	Low	Moderate	Moderate	Moderate	High	High
Amount DNA required	High	Low	Moderate	Low	Low	Low
DNA quality	High	High	Moderate	Low	Low	Moderate
Cost per analysis	High	Low	Moderate	Low	Low	Low

technique available capable of differentiating GM transformation-events. The increasing insight provided by the genomics era presents wider possibilities to compare gene structure and function in divergent organisms. *Comparative mapping* allows the transfer of information among orthologous genes or homologous chromosomes. This is not only useful for better mapping, gene cloning and characterization but also for marker discovery (Sorrells and William 1997) and the comprehension of the processes' underlying genetics.

6.2.1 *Molecular Approaches Used in Practical Plant Breeding*

Nonetheless, when it comes to practical plant breeding, the use of these molecular applications is usually limited. Breeders' approach is very pragmatic, for a character under study the benefits of using molecular tools have to be greater than the assay-cost. There are however two approaches where marker technology are undeniably very valuable.

First, as it enables DNA markers to be linked to traits of interest (Fig. 6.1) and to direct the selection towards these markers instead of the phenotypic reaction of superior plants (Edwards and Mogg 2001). Hence, the selection of desirable genotypes can be done directly at the DNA-level in a non-destructive manner with no interference of the environment and regardless of the plant developmental stages, thus allowing a greater efficiency of field trials (Peleman and van der Voort 2003). The use of molecular information can enhance breeding strategies based on phenotype-selection, which is broadly referred to as marker-assisted selection (MAS, Dekkers and Hospital 2002). In practice, markers rather than known genes, are likely to be used (Villanueva et al. 2002). Added value can be created through the introduction of new traits that would have been difficult or required additional

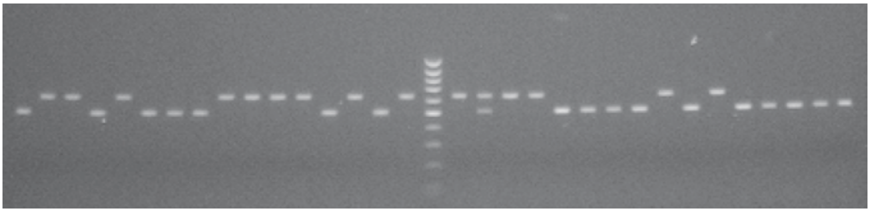


Fig. 6.1 Example of a co-dominant marker useful for MAS and for bulk analysis of advanced breeding lines. The PCR marker from F Rath allows the detection of barley individuals with a wanted/unwanted 'final attenuation' profile for malting barley as well as individuals segregating for this trait

steps by classical breeding e.g. difficulties in phenotypic scoring, selection of rare recombinants or necessity of test crossing (Dayteg et al. 2008; Tuvesson et al. 1998). As several markers can be used for selection, new possibilities to incorporate different genes into the same line are given to the plant breeders, e.g. attempting to slow down the evolution of pathogen virulence (Hospital 2003; Werner et al. 2005). Knapp (1998) showed, in his models, that a breeder using phenotypic selection must test 1.0 to 16.7 times more progeny than a breeder using MAS to be assured of selecting one or more superior genotypes. However, the advantages of MAS over phenotypic selection are considerably reduced when conducted in later generation (Liu and Knapp 1990). Consequently MAS though providing more accurate responses also dramatically increases the frequencies of superior genotypes in early generations.

Then, modern plant breeding is not only based on genotype-building but also on manipulating variation within gene-pools of a cross. DNA-fingerprinting of breeding lines using molecular markers (Fig. 6.2), as well as detailed genome analysis of plants, provides in this aspect a very powerful and efficient tool to characterize, monitor and protect germplasms (Lombard et al. 2000). Multilocus marker-types are usually preferred for their discrimination potential, as they reveal polymorphism information at several loci simultaneously. However, any set of representative DNA-markers is capable to cover the whole genome (Gupta et al. 1999). A broad genetic diversity is paramount in e.g. resistance breeding because of the rapid evolution and occurrence of new and virulent races of pathogens. Unfortunately, most elite cultivars, especially small-grain cereals, are breed on a quite narrow genetic base, and the limited genetic diversity may impede the deployment of new sources of resistance for a pathogen or the discovery of new positive alleles for a character. The introduction of novel characteristics from unadapted wild or exotic germplasms into elite breeding lines have shown to counteract this limitation (Ivandic et al. 1998; Ordon et al. 1996). Wild relatives or progenitors can represent a rich source of useful resistance genes, as *H. vulgare* ssp. *spontaneum* is for barley in regards to new sources of resistance to leaf rust, powdery mildew, barley yellow dwarf virus, scald, net blotch, *septoria*, etc. (Fetch et al. 2003; Jahoor and Fischbeck 1987).

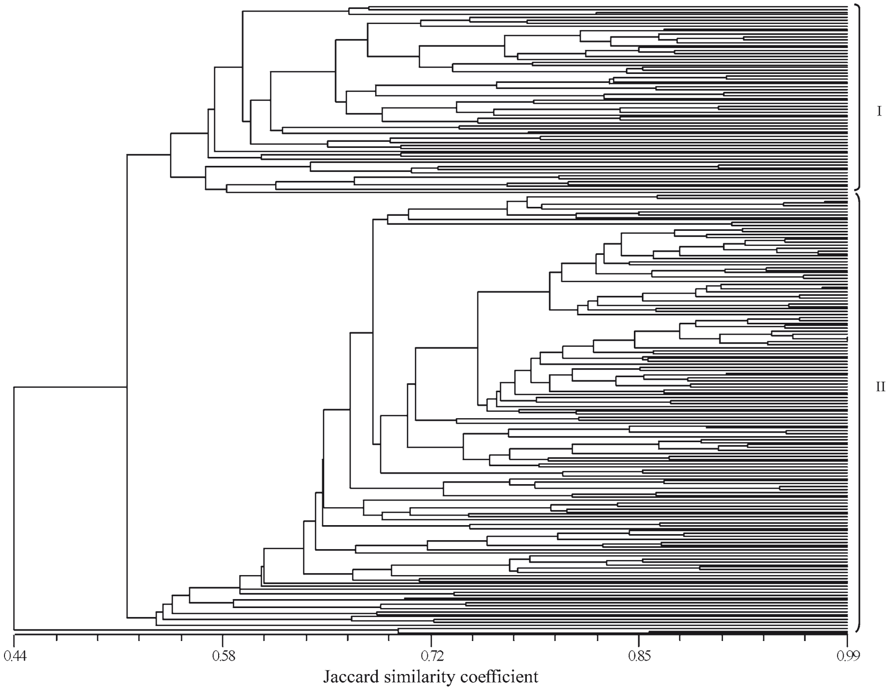


Fig. 6.2 Example of an UPGMA dendrogram illustrating the genetic relationship between 227 Nordic and Baltic barley accessions. Source: Kolodinska-Brantestam et al. (2004). Molecular markers enable genotype-discrimination and the estimation of genetic relatedness between lines for an effective exploitation of the germplasm (see text)

Breeders, however, are usually reluctant about using wild germplasms in their breeding programs because of complex, long-term and unpredictable outcomes, particularly in crops where quality traits are important criteria (Peleman and van der Voort 2003). Marker assisted backcrossing (MAB) is an effective aid to selection in backcrossing: first as the target trait can be directly monitored, hence avoiding phenotypic scoring. Then, as markers closely linked to the target gene can limit the surrounding DNA from the donor parent, thus removing possible linkage drag. Finally, as markers dispersed over the genome permit the selection of progeny with higher proportions of the recurrent parent genetic background (Holland 2004). Any kind of DNA markers can be used, however codominant markers are considered to be the most useful as they allow the selection of heterozygous individuals, as Chen and colleagues (2000) have shown using 128 RFLP loci to MAB of the *Xa21* gene in rice. Melchinger (1990) reviewed the advances of MAB. He compared conventional schemes described by Allard (1960) to MAB models described by Tanksley and Rick (1980). They demonstrated that the proportion of the recurrent parent in the first generation of MAB could correspond to that expected after three generations

of conventional backcrosses. These results were verified by Frisch and colleagues (1998) who estimated to two the number of generation needed to obtain a genotype with 98% or 99% genetic similarity to the recurrent parent. Considering that Allard estimated the adequate number of generations to six, MAB represents a considerable gain of time. However, they also stressed that the number of markers and material to be screened would be very large.

6.2.2 Need for Molecular Marker Automation

As previously described, breeding resources can be efficiency exploited using molecular markers first, by reducing the number of inadequate lines requiring extensive phenotypic evaluation in later generations (Holland 2004) and then by optimising the use of the gene-pools and finally, by speeding up the introgression process of new characters. In a practical breeding perspective, however, this requires an adaptation of the methodology to allow plant material to be monitored in realistic high number of individuals in early generations (Dayteg et al. 2007). The high number of individuals and the economic constrains involved in a breeding program compel molecular markers to be technically easy to use, cheap and informative (Hernandez 2004). While most PCR-based markers fulfil these requirements (Table 6.1), automating PCR-procedures faces a few problems. First, amongst the PCR-based markers, there is not today a single established or universal marker technology and each type of marker might require its own procedure. Then, marker technology as a whole is in a growing phase and evolves rapidly. Technologies as well as the availability of the appropriated markers may constitute a shortage in the practical approach to marker applications. Finally, DNA-marker being a broad concept, each of its specific application might require its own marker technology or technical challenges (Dayteg et al. 2007).

The evolution of robotics in biotechnology and the progress of bioinformatics have been significant for the development of high throughput system (Cahill and Schmidt 2004). However, the spin-off effects of the pharmaceutical industry remain limited in practical breeding due to their important investment costs. Because of their economical value major crops have essentially been in focus for such investments.

6.3 Experience of Automation at Svalöf Weibull Laboratory

Though molecular markers are today well established practice in plant breeding the automation of the technology is still in its cradle. Ready-made robotic applications can easily provide some answers to specific issues, i.e. sample extraction, sample preparation etc. (Fig. 6.3), if the investments are possible. However, we found these equipments often to be developed according to very specific protocols or routinely

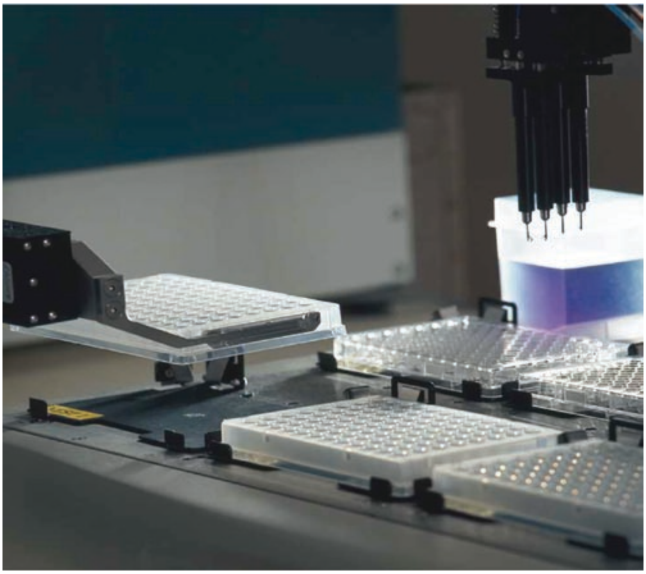


Fig. 6.3 Automation of pipetting procedures is the first obvious moment in regards to improving work conditions and throughput

adapted to special commercial kits. They present, therefore, poor flexibility to already established in-house practices and more generally to the processes and economical constraints of practical breeding. There is not any ready-to-use automation solution. The concept of automation or high throughput (HT) in themselves remain largely in "the eye of the beholder" e.g. an increase from 50 to 100 DNA-extractions per day might be considered HT for some laboratories but insignificant for others. Therefore, in a more general manner HT should rather be seen as an appreciable increase of the productivity (in percent) and automation as an attempt to decrease manual labour from standardized workflows. This chapter does not claim to hold the ultimate key to automated HT applications in plant breeding but to simply lay down the principles used in a very practical approach which might be found useful for others.

It is primordial for plant breeding companies to keep focus on their main activities; it seemed therefore, more justified to adapt the molecular processes to the breeding programmes than *vice-versa*. Because of the cost involved in automation, it is of great importance to really understand the molecular needs and requirements necessary to achieve the goals set by modern plant breeding, and to carefully analyse the methodology for maintaining enough flexibility to be able to adapt to its challenges.

The whole molecular workflow was therefore first subjected to an "intellectual exercise" and the automation-possibilities were evaluated in a three step procedure as schematised in Fig. 6.4. In the *analystic stage* the current state of the workflow is established and the prospective state characterised in terms of usage

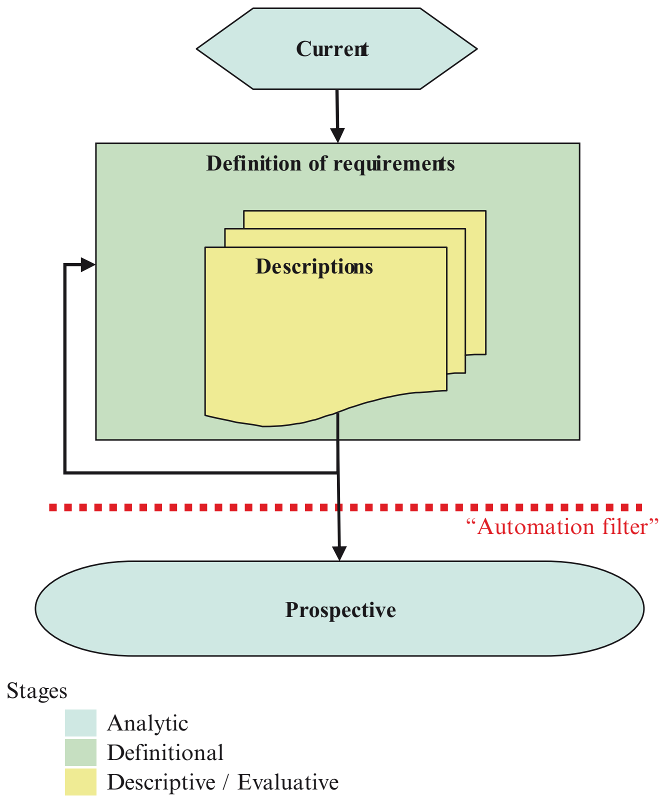


Fig. 6.4 Automation exercise. Mental process composed of three stages evaluating the automation possibilities of molecular workflows. The analytic stage establishes the current state of the workflow and characterizes the prospective state. The definitional stage defines the requirements needed to achieve the prospected goal. The descriptive stage details each of the defined steps and evaluates for each one visible bottlenecks and possibilities of improvement. These "improvements" are then subjected to the "automation filter". In that step the automation-feasibility is tested for each one of them and depending of the results the step can either be accepted in the prospective goal or redefined

of the molecular tools (i.e. applications required in breeding programmes), identification of necessary molecular tools and expectation of the laboratory's capacity. The requirements needed to achieve this prospective goal, i.e. all the required procedures in the process are defined in the *definitional stage*. They are then detailed into operation-steps in the *descriptive stage*. Lydiate (1999) has described an efficient genomic research as a three steps procedure (1) automating what can be automated (2) speeding up the process (3) allowing molecular shortcuts. A similar approach was applied at this stage to identify all possible bottlenecks and to define possibilities of improvement at each step and subject them to an "automation-filter". This simply means that each of them are tested for their automation-ability, which is to evaluate if automation is feasible for this specific step in terms of robotic

availability, staff skill and accessibility, cost/gain evaluation and if an eventual automation could present new bottlenecks (i.e. extra procedures). This final evaluation is necessary to either redefine or accept the improved procedure (with or without automation).

6.3.1 Characterisation of Breeding Activities

In regard to commercial plant breeding activities, we established that the application of molecular markers can be characterised in the two main groups described below and summarised in Fig. 6.5:

MAS, shifting the traditional phenotype-based selection to genotype-based selection, is routinely used in plant breeding programs mainly for selecting alleles with large effects on traits with relatively simple inheritance (Holland 2004). The technology empowers breeding programmes, on its own or in combination with phenotypic testing, when selecting for:

- Traits with small phenotypic effects i.e. when the phenotype is a poor predictor of the breeding value (low heritability).
- Traits difficult or expensive to assess (e.g. nematode resistance, Barley Yellow Dwarf Virus-resistance).
- Plants heterozygous for recessive traits (e.g. powdery mildew *ml-o* resistance in barley requires one more generation).

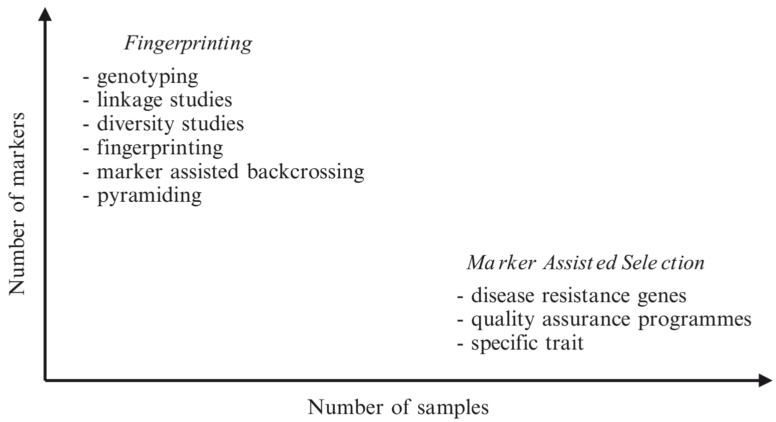


Fig. 6.5 Characterisation of DNA-marker projects in practical plant breeding programmes. The relationship between the number of markers/number of individuals for the type of application is used as the base of the representation. Adapted from Dayteg et al. (2007). The ranking within groups has been made arbitrarily and may not be representative as the figures vary between crops and studies

- Traits expressed in a late development stage or where the individual needs to be sacrificed to score its phenotype (e.g. male sterility in *Brassica napus*, final attenuation in malting barley).
- Alleles not expressed in the selection environment.
- Combining traits that might mask each other's effects (e.g. pyramiding resistance genes).
- The technology is indispensable for GM-quality control of commercial cultivars.

This kind of applications usually implies a limited amount of molecular markers used to screen large number of samples.

Whole genome study, or ***fingerprinting***, enables the characterization of genotypes and the estimation of genetic relatedness between lines. This information is crucial to allow plant breeders to appropriately choose the parental lines for their crosses especially for hybrid production (Ma et al. 2003), but also for an effective exploitation of the germplasm by monitoring the diversity of their gene-pools (Kolodinska-Brantestam et al. 2006). The use of molecular markers for genetic studies have been much diverse, the main applications include:

- Identification and fingerprinting of genotypes
- Assessment of genetic variability and/or line purity (e.g. conservation or expansion of the gene-pool, pure line or inbreeds-check)
- Estimation of genetic relatedness between breeding material and/or populations (e.g. estimate of heterosis, allele frequency)
- Foreground (genotyping at target loci) and background (genotyping at loci across the genome) selection for marker assisted backcrosses (MAB) (e.g. introgression of novel traits from unadapted germplasm into elite breeding lines)
- Increase of the genetic variability of improved lines (single large-scale marker-assisted selection (SLS-MAS), Ribaut and Bertrán 1999)
- Characterisation and rare allele selection of exotic germplasm
- Linkage analysis

This kind of applications usually implies a rather restricted number of samples screened with a large number of markers.

6.3.2 Characterisation of Molecular Activities

Considering these two kind of applications, and the relative moderate scale of the SW laboratory's molecular operations, the choice of marker technologies was limited by focusing exclusively on PCR-based markers. Other groups may opt for different technical solutions to fulfil their applications and/or scale, but the principle remains the same. In SW's case, PCR-based markers were chosen because, as seen in Table 6.1, they fulfil most of the requirements necessary in practical plant breeding. They are easy to use, require small amount of crudely extracted DNA, enable automation and are relatively cheap. Within PCR-based

markers, microsatellites (SSRs) are especially interesting as they are well spread on the genome, generally highly informative, widely available and well described in most of the crops. Their ease of detection via automated-systems makes them currently the most popular PCR-based marker in cereal breeding (Korzun 2003) and their flexibility allows their application in the two main groups described above.

Focus was given to increasing the **throughput** and lowering the **cost** for the two groups and two specific sets of requirement could be defined and highlighted in Fig. 6.6. In *MAS*-like applications, focus was given to decreasing procedure time for individual sample thus allowing a larger number of samples to be processed for a given time period. As few datapoints are gathered for each sample, neither DNA quality nor pipetting accuracy are a major issue and can therefore be simplified to their quickest, though reliable, form. Because the assays are well described data-acquisition can be focused and only limited to the expected fragments for more efficiency. In the case of *Fingerprinting*-type applications, focus was given to increasing the sample efficiency e.g. by multiplexing (simultaneous amplification of several molecular markers in a single reaction). This sets higher requirements on DNA quality, automated liquid handling and data-analysis (Mace et al. 2003).

6.3.3 Automation of Analytic-processes

The whole marker analyse-process was decomposed in a few components and each subjected to the approach (Fig. 6.4) in order to improve analysis capacity from tens of thousands to hundreds of thousands. All processes were standardized by working solely on microtiter-plate format from start to finish.

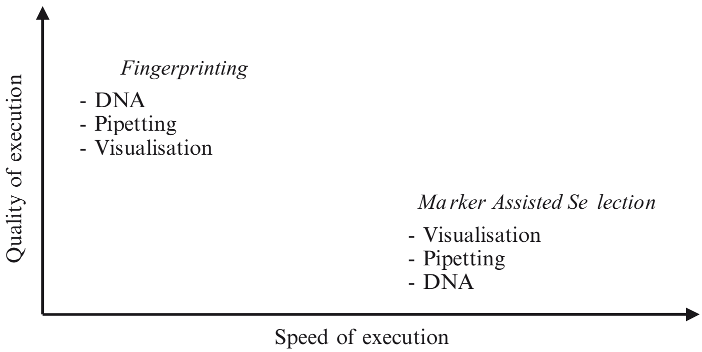


Fig. 6.6 Requirements for the automation of the main application groups. Increased capacity is obtained by emphasising on **speed** for MAS and **multiplexing** for Fingerprinting. Adapted from Dayteg et al. (2007)

6.3.3.1 Sampling and DNA Processing

Plant samples are collected in the field or in greenhouses using a paper punch devise and placed to the appropriate position in 96-well plates. Plates are kept cool during the collection process. Two people can “harvest” up to 3,840 leaf samples per day. When handling GM-material special precautions, such as sterilizing the paper punches to avoid possible cross-contamination, must be taken into consideration, which will reduce the harvesting capacity by one quarter. Once collected, improvements to the in-lab procedures allow a rapid and efficient DNA-processing. The DNA isolation is generally performed according to a quick DNA-extraction protocol (Dayteg et al. 1998) (Fig. 6.7) enabling the DNA to be processed within 10 min (theoretically more than 4,000 samples in a working day). For methods requiring larger DNA-quantity of better quality a “quick standard” method has been adapted from Cheung et al. (1993) enabling the extraction of ca 800 samples per day, by processing the samples-plates in a robotic grinder devise. When handling samples from remote locations a seed-based DNA extraction protocol, as described by von Post et al. (2003), can be used to rapidly extract large amount of material (700 samples per day and person). The automation of these protocols has been successfully tested, on the system described below. Nonetheless, because they impeded the accessibility of the system to other, more demanding, procedures they remained principally manual. Parallel robotic equipment has been proposed as a possible solution. However, the marginal gain of time, and/or capacity, does not really justify the investment costs.

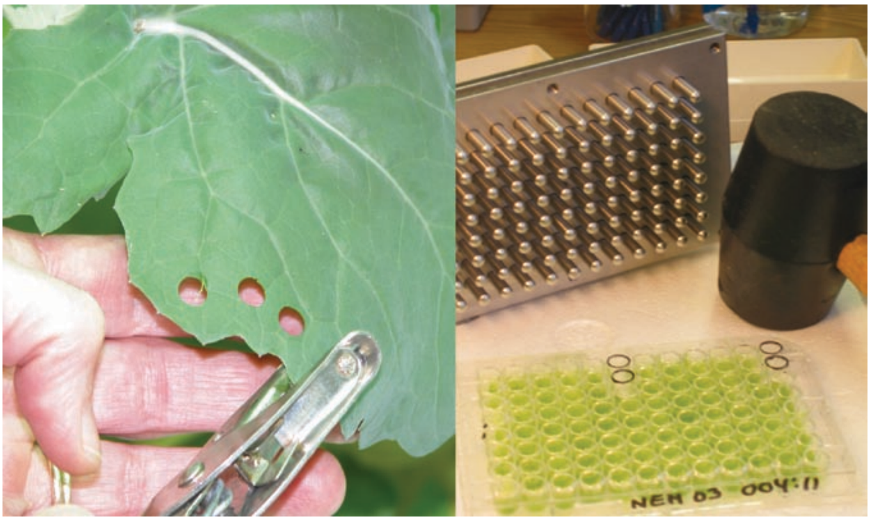
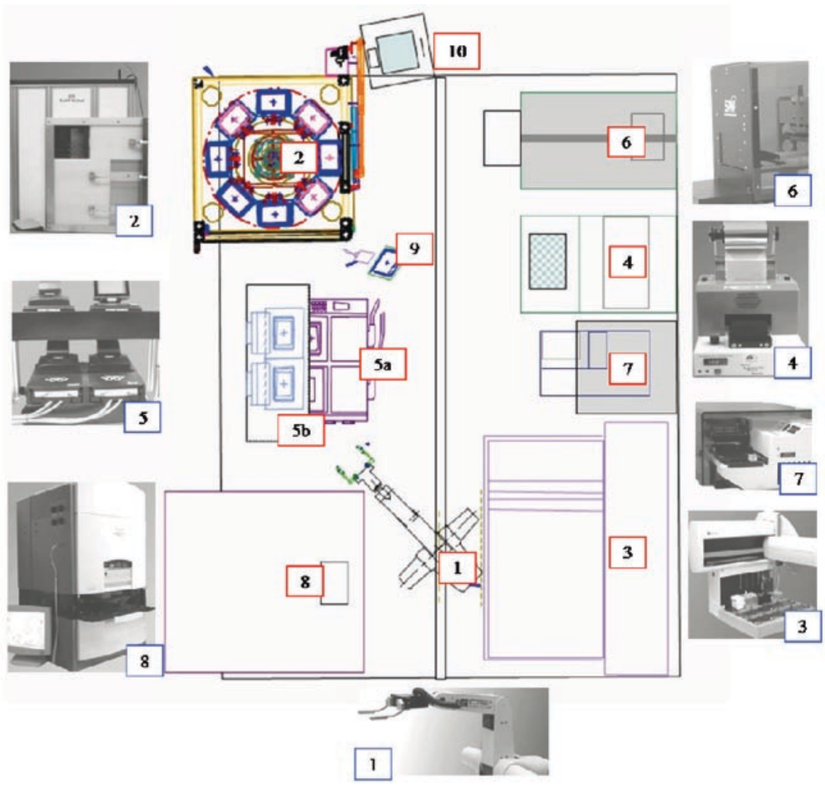


Fig. 6.7 Pictures of a sampling-punch and of a collection plate after a quick DNA-extraction. This simple procedure efficiently enable the crude extraction of thousand of DNA-samples in a working day for PCR-based molecular assays (Dayteg et al. 1998)

6.3.3.2 PCR Procedure and Data Acquisition

Development of robotics for molecular analyses has been essential. In collaboration with Thermo CRS (Burlington, Canada), a fully automated system was developed with the main emphasis on flexibility and high throughput. The system, constituted of different peripherals, is served by a robotic arm as described in Fig. 6.8. The components have not only been chosen for their individual automated performances but also because they all feature an open architecture that allows their nests to be



1	CRS robotic arm on 2m track	Transporting and serving plates to the other peripherals
2	CRS Refrigerated carousel	Housing 120 plates for cold storage (+4°C)
3	Packard *Multiprobe II*	Automated pipetting robot w/ 8 accessible nests
4	ABGene *APLS100*	Sealing of PCR plates
5a	MJR *Tetrad PTC-225*	Thermocycler with 4 independent PCR-blocks and automated lids
5b	MJR *Remote Connectors*	Allows 2 PCR-block to be placed as satellites (for nest accessibility)
6	CRS Piercer	Unsealing PCR plates after amplification.
7	Thermolab *Multidrop 384*	Automated dispensing robot for rapid mix dispensing. Placed in a fume hood.
8	SpectruMedix SC*E 9610*	Autom.ated 96 capillaries electrophoresis w/ 6 accessible nests
9	Barcode reader	Registration of plates barcode
10	Air-tight Waste	Disposal of plates containing formamide. Connected to fume hood.

Fig. 6.8 Illustrated blueprint of the fully automated molecular marker assay system developed at SW laboratory. Adapted from Dayteg et al. (2007)

accessible by the robotic arm, thus enabling full automation. Thermo CRS has supervised the integration of these peripherals into **one** core system.

The two main application groups described above are accomplished using a common core set-up. The robot arm handles sample-plates as such: samples are moved to a pipetting devise for PCR-setup. The reaction volume varies between 5 and 25μL. A volume of 1.5 μL DNA is transferred to each of the corresponding PCR-wells. Between each transfer the tips of the pipetting robot are thoroughly washed to avoid cross-contamination. While DNA-plates are moved back into cold-storage, the PCR-plates are sealed to avoid evaporation and placed in the thermocycler where the appropriated PCR is performed. They are then placed in cold-storage while waiting for visualisation. Prior to visualization, the PCR-plates are pierced. The samples are then visualised on the 96-capillaries electrophoresis, either directly with ethidium bromide or first denatured with formamide containing a fluorescent internal line standard.

The use of a core system, besides its flexibility, increases the reliability of the system, reduces start-up time and enables different assays to be run simultaneously (Brandt 1998). Furthermore it enables the whole process to be performed in a fully automated manner hence freeing such procedures from any human interventions, since it often is the major cost, a decrease in "hands-on" will also decrease the cost (Klapper et al. 1998). The fully automated nature of the system increases further the assay-capacity through the possibility of extended overnight and week-end runs.

6.3.3.3 Data Handling and Sample Tracking in Mass Number

The use of capillary electrophoresis increases not only the sample throughput but enhances the detection-sensitivity. The development of specific softwares to handle electrophoresis-data has allowed the use of ethidium bromide (Fig. 6.9) and fluorescently labels for direct visualization of DNA-fragments. Non-fluorescently dyed amplicons are directly visualized in an ethidium bromide gel-matrix. This beneficently replaces agarose-gel based analyses of trait-linked markers. This type study is especially applicable for MAS, for quality assurance programmes to monitor the adventitious presence of GMOs in commercial varieties or even pyramiding several marker-linked traits. Fluorescently-dyed amplicons allow advantageous multiplex reactions and replace polyacrylamide-gel based electrophoresis. This type of study is especially applicable for whole genome studies e.g. MAB, genetic relationship studies etc.

The capillary electrophoresis decreases "hands-on" time, and its enhanced sensitivity enables a decrease up to 1/5th of the reaction volume, hence significantly lowering assay-costs. Additionally, it enables to speed up both data-acquisition and analysis time, hence increasing the assay-capacity.

In order to cope with this mass numbers of data, the laboratory has entered the next phase of development by integrating a sample tracking system to improve the sample-monitoring from tissue arrival to data export which will allow an improved traceability and a stronger breeder-laboratory interface.

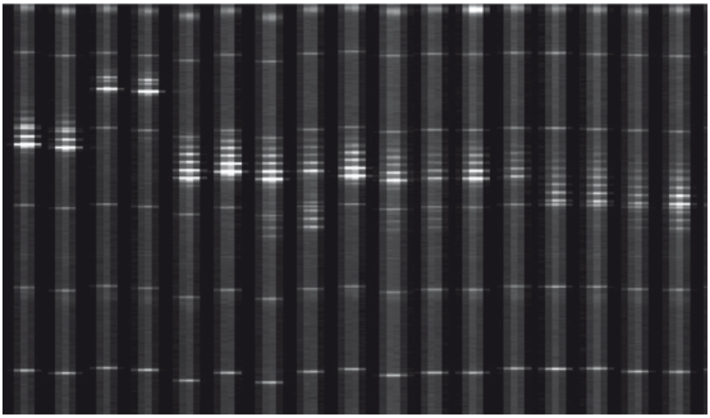


Fig. 6.9 Example of a SSR amplification visualised on 96 capillary electrophoresis with an ethidium bromide-containing matrix

6.3.4 Automation Performance

To evaluate the system's performance, cost and throughput of a typical assay visualized with an agarose-gel based electrophoresis was compared with similar ones performed on the automated-system (Table 6.2). The two visualisation modes were used, i.e. non-fluorescent amplicons visualised with ethidium bromide and fluorescently dyed amplicons. The first striking result, as seen in the last column, is the extraction cost. This is due to the more demanding DNA quality used in fluorescent assays. Because of a more time-consuming DNA extraction protocol and data-analysis the hands-on time is greatly increased. The PCR cost is also slightly higher than with ethidium bromide-stained capillary-electrophoresis. This is due to a larger PCR-volume and the use of fluorescent dyes. The analysis cost is, however, much higher due to more expensive gel-matrix and the use of Internal Lane Standards. However, it is believed that a further reduction in reaction volume might be possible and it should be noted that the use of multiplex reactions, though slightly increasing the PCR cost, multiply the throughput. That means that from the noted capacity of ca 900 datapoints (dp)/day at an assay cost of 1.15€/sample, the use of four multiplexed markers with an automated procedure, increases this figure to more than 3,600 dp/day for nearly the same overall sample price. In regards to the automated assay visualised with ethidium bromide the overall cost is lowered and the throughput largely increased. The lower reaction volumes, the increased capacity and the minimal "hands-on" time efficiently counteract the investment costs. The assay-throughput, depending on the application, ranges from ca 11,000 dp/week to more than 24,000 dp/week (using four multiplexed markers).

The developed system enables the automation of PCR-based markers and provides a competitive platform for their large-scale use in plant breeding. Automated systems, such as the one presented, can even enhance the detection of linked markers.

Table 6.2 Comparison of amplicons visualization-methods. The costs (in €) are calculated per plate (96 samples) at their respective capacity. All figures are given for simple reaction (simplex). The extraction cost includes the material and consumable costs. The first two columns are for the quick-extraction protocol (4,000 samples capacity), while the last is for the "quick standard"-protocol (800 sample capacity). The PCR cost includes the material and consumable costs for the respective applications. The two last columns do not include the cost of tips as they are performed on the robot. Furthermore, their PCR-volumes have been decreased to 1/5th and 1/3rd respectively. The analysis cost includes the material (e.g. ladder, capillaries) and consumable costs. There are two figures for the last column, as the use of commercial or home-made ILS strongly influences the cost. Labour is calculated as the average time. Differences in the depreciation cost are due to the cost of the respective equipment. Adapted from Dayteg et al. (2007)

Electrophoresis	Agarose based	Capillary based	
	Ethidium bromide	Ethidium bromide	Fluorescent dyes
Extraction cost	1.5	1.5	82
PCR cost	20.5	6.2	8.3
Analysis cost	5.6	6	18.2[a]/10.3[b]
Depreciation[c]	0.9	9.3	9.3
Hands-on time (average)	55 min	16 min	50 min
Throughput (per day)	900 samples	2,200 samples	900 samples

[a]Use of commercial ILS.
[b]Use of non commercial ILS.
[c]On a 10 year basis and 200,000 analyses per year.

Nonetheless, today the rather limited amount of molecular markers, amenable to automation and diagnostically linked to genes of practical breeding interest, remains the major drawback of such application. Confidently, the current and rapid developments in the molecular field (e.g. association mapping) will provide an appreciable quantity of molecular tools, adapted to practical plant breeding, that will increase the diversity of automated assays.

6.4 Conclusion

Plant breeders act in a very competitive market, continuously re-evaluating their products in an attempt to respond to ever-changing market demands and agricultural practices. The undeniable benefits of molecular markers and the constant decrease of PCR-costs provide considerable allies in such never-ending "selection-pressure". For these reasons, PCR-markers are considered a valuable tool in the most various, and even modest, breeding programmes. However, molecular technology is still often viewed as an additional cost in many breeding procedures. An improvement of the productivity, through procedure-optimizations and/or automation, will provide a further decrease of the assays cost and increase its availability. To date the costs for automating marker technologies in applied plant breeding still

represent an important investment. Nonetheless, it cannot be stressed enough, that automation is the ultimate steps in a series of optimizations of molecular (and breeding) procedures and that the automation concept itself very much depends of the beholder's perspectives. For our part, the main goal was to decrease to the maximum any human-interactions from tedious, unqualified, and non-rewarding, procedures thus liberating the highly qualified staff to perform more complex and demanding applications. The fully-automated system developed at SW, like any automated molecular screening settings, provides the ability to efficiently generate large dataset from either a large amount of material and/or markers. This system is flexible enough to adapt to the requirements of practical breeding in regards to molecular analyses i.e. to assay realistically high numbers of samples as promptly and as cheaply as possible and thus enabling practical "molecular breeding".

It is important to keep in mind that automation of "molecular breeding" is an ongoing process, not only in terms of technical development but rather as a constant questioning of the different breeding-specific applications, trying to fulfil the main goal of breeders: to promptly release the best product quality at the lowest cost. The increasing amount of sequence information and the determination of gene function are leading to the use of emerging marker types such as SNPs. They hold great promises of rapid and highly automated genotyping (Gupta et al. 1999; Korzun 2003).

Acknowledgment Formas and Øforsk are acknowledged for their financial support.

References

Allard RW (1960) Principles of Plant Breeding. John Wiley, New York

Bergal P, Friedberg L (1940) Essai d'identification des orges cultivées en France. Ann Epiphyt phytogenet VI:2–4

Brandt DW (1998) Core system model: understanding the impact of reliability on high-throughput screening systems. Drug Discov Today 3(2):61–68

Cahill DJ, Schmidt DH (2004) Use of marker assisted selection in a product development breeding program. In: New directions for a diverse planet. Proceeding of the 4th International Crop Science Congress. Brisbane, Australia

Chen S, Lin XH, Xu C et al (2000) Improvement of bacterial blight resistance 'Minghui 63' an elite restorer line of hybrid rice, by molecular marker-assisted selection. Crop Sci 40:239–244

Cheung WY, Hubert N, Landry BS (1993) A simple and rapid DNA microextraction method for plant, animal and insect suitable for RAPD and other PCR analyses. PCR Meth Appl 3:69–70

Dayteg C (2008) Automation of molecular markers in practical breeding of spring barley (*Hordeum vulgare* L.). ISBN: 978-91-85913-31-2

Dayteg C, Rasmussen M, Tuvesson S et al (2008) Development of an ISSR-derived PCR marker linked to nematode resistance (*Ha2*) in spring barley. Plant Breed 127:24–27

Dayteg C, Tuvesson S, Merker A et al (2007) Automation of DNA marker analysis for molecular breeding in crops: practical experience of a plant breeding company. Plant Breed 126:410–415

Dayteg C, von Post L, Öhlund R et al (1998) Quick DNA extraction method for practical plant breeding programmes. In: Plant and animal genome VI. San Diego, CA

Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. Nature 3:22–32

Edwards KJ, Mogg R (2001) Plant genotyping by analysis of single nucleotide polymorphisms. In: Henry RJ (ed) Plant genotyping: the DNA fingerprinting of plants. CABI, UK
Fetch TGJ, Steffenson BJ, Nevo E (2003) Diversity and sources of multiple disease resistance in *Hordeum spontaneum*. Plant Dis 12:1439–1448
Frisch M, Bohn M, Melchinger AE (1998) Comparison of selection strategies for marker-assisted backcrossing of a gene. Crop Sci 39:1295–1301
Gupta PK, Varshney RK, Sharma PC et al (1999) Molecular markers and their applications in wheat breeding. Plant Breed 118:369–390
Helentjaris TG, King G, Slocum M et al (1985) Restriction fragment length polymorphism as probes for plant diversity and their developments as tools for applied plant breeding. Plant Mol Biol 5:109–118
Hernandez P (2004) Development and use of genomic tools for cereal introversion breeding. In: Vollmann J, Grausgruber H, Ruckenbauer P (eds) Genetic variation for plant breeding. Proceedings of the 17th EUCARPIA General Congress. BOKU, Vienna
Holland JB (2004) Implementation of molecular markers for quantitative traits in breeding programs – challenges and opportunities. In: New directions for a diverse planet. Proceedings of the 4th International Crop Science Congress. Brisbane, Australia
Hospital F (2003) Marker-assisted breeding. In: Newbury HJ (ed) Plant molecular breeding. Blackwell, UK
Ivandic V, Walther U, Graner A (1998) Molecular mapping of a new gene in wild barley conferring complete resistance to leaf rust (*Puccina hordei* Otth). Theor Appl Genet 97:1235–1239
Jahoor A, Fischbeck G (1987) Sources of resistance to powdery mildew in barley lines derived from *Hordeum spontaneum* collected in Israel. Plant Breed 99:274–281
Klapper PE, Jungkind DL, Ferner T et al (1998) Multicenter international work flow study of an automated polymerase chain reaction instrument. Clin Chem 44:1737–1739
Knapp SJ (1998) Marker-assisted selection as a strategy for increasing the probability of selecting superior genotypes. Crop Sci 38:1164–1174
Koebner R (2003) MAS in cereals: Green for maize, amber for rice, still red for wheat and barley. In: Marker assisted selection: A fast track to increase genetic gain in plant and animal breeding? Session I: MAS in plants. Proceedings of an international workshop organised by the Fondazione per le Biotecnologie, the University of Turin and FAO. Turin, Italy
Kolodinska-Brantestam A, von Bothmer R, Dayteg C et al (2004) Inter simple sequence analysis of genetic diversity and relationships in cultivated barley of Nordic and Baltic origin. Hereditas 141:186–192
Kolodinska-Brantestam A, von Bothmer R, Rashal I et al (2006) Genetic diversity changes and relationships in spring barley germplasm of Nordic and Baltic areas as shown by SSR markers. Genet Resour Crop Ev 54:749–758
Korzun V (2003) Molecular markers and their applications in cereals breeding. In Marker assisted selection: A fast track to increase genetic gain in plant and animal breeding? Session I: MAS in plants. Proceedings of an international workshop organised by the Fondazione per le Biotecnologie, the University of Turin and FAO. Turin, Italy
Liu BH, Knapp SJ (1990) GMENDEL: A program for Mendelian segregation and linkage analysis of individual and multiple progeny populations using log-likelihood ratios. J Hered 81:407
Lombard V, Baril CP, Dubreuil P et al (2000) Genetic relationships and fingerprinting of rapeseed cultivars by AFLP: Consequences for varietal registration. Crop Sci 40:1417–1425
Lydiate D (1999) Saskatoon genomics initiative could revolutionize canola research. The Agbiotech bulletin
Ma C, Fu T, Bengtsson L et al (2003) Genetic diversity of Chinese and Swedish *Brassica napus* cultivars based on Inter-SSR PCR markers and its relationship to hybrid performance. J Swed Seed Assoc 2:67–77
Mace ES, Buhariwalla HK, Crouch JH (2003) A high-throughput DNA extraction protocol for tropical molecular breeding programs. Plant Mol Biol Rep 21:459a–459h

Melchinger AE (1990) Use of molecular markers in breeding for oligogenic disease resistance. Plant Breed 104:1–19

Nilsson-Ehle H (1914) Zur Kentnis der mit der Keimungsphysiologie des Weizens in Zusammenhang stehenden innerern Faktoren. Z Pflanzenzuecht 2:153–157

Ordon F, Weyen J, Korell M et al (1996) Exotic barley germplasms in breeding for resistance to soil-borne viruses. Euphytica 92:275–280

Peleman JD, van der Voort JR (2003) The challenges in marker assisted breeding. In: van Hintum ThJL, Lebeda A, Pink DA, Schut JW (eds) Eucarpia leafy vegetables. CGN, Wageningen

Ribaut J-M, Bertrán J (1999) Single large-scale marker-assisted selection (SLS-MAS). Mol Breed 5:531–541

Saiki RK, Gelfand DH, Stoffel S et al (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Sorrells ME, William WA (1997) Direct classification and selection of superior alleles for crop improvement. Crop Sci 37:691–697

Tanksley SD, Rick CM (1980) Isozymic gene linkage map of the tomato: applications in genetics and breeding. Theor Appl Genet 57:161–170

Tuvesson S, Dayteg C, Hagberg P et al (2007) Molecular markers and doubled haploids in European plant breeding. Euphytica 158:278–294

Tuvesson S, von Post L, Öhlund R et al (1998) Molecular breeding for the BaMMV/BaYMV resistance gene *ym4* in winter barley. Plant Breed 117:19–22

Werner K, Friedt W, Ordon F (2005) Strategies for pyramiding resistance genes against the barley yellow mosaic virus complex (BaMMV, BaYMV, BaYMV-2). Mol Breed 16:45–55

Villanueva B, Pong-Wong R, Williams JA (2002) Marker assisted selection with optimised contributions of the candidates to selection. Gen Sel Evol 34:679–703

von Post R, von Post L, Dayteg C et al (2003) A high-throughput DNA extraction method for barley seed. Euphytica 130:255–260

Chapter 7
Pyramiding Genes for Enhancing Tolerance to Abiotic and Biotic Stresses

Raveendran Muthurajan and Ponnuswami Balasubramanian

Abstract Biotic stresses and abiotic stress factors such as salt, drought, cold and extreme temperatures severely limit crop productivity. Attempts to improve crop varieties for traits conferring tolerance to these stresses are being made continuously to sustain the food production. Molecular breeding and genetic engineering strategies are serving as efficient tools to accelerate the process of developing stress tolerant genotypes. Developing crop varieties that can withstand incidence of multiple stresses is one of the major breeding objectives nowadays, wherein series of target genes identified in different parents are accumulated into a single genotype (gene pyramiding). Generation of transgenic plants by introducing two or three foreign genes is one of the common means to develop plants exhibiting resistance against multiple stresses. Recently there have been advances in the mapping of genes controlling quantitative traits, through quantitative trait loci (QTL) mapping experiments and analysis of genomic data. Such advancements in genome mapping and high-throughput laboratory genotyping protocols have enabled us to monitor introgression of desired genes/regions from two or three different sources into a single background. In this paper, we have attempted to review the achievements made in the field of generation of improved crop varieties by pyramiding (genetic transformation and marker assisted gene pyramiding) desired genes from different sources.

7.1 Introduction

World's crop production is facing serious threats from the occurrence of biotic and abiotic stresses which contributes around 60% of production loss during the cropping period. Despite concerted efforts through breeding programs aimed at developing biotic and abiotic stress tolerant crop varieties by exploiting endogenous

R. Muthurajan and P. Balasubramanian (✉)
Centre for Plant Molecular Biology, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu 641 003, India
e-mail: balasubrap@hotmail.com

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_7, © Springer Science+Business Media B.V. 2010

resistance genes, the progress is very limited because of the complex nature of resistance mechanism(s). Large amount of chemical pesticides and fungicides are being used which lead to increase in cost of production and in addition it causes environmental pollution.

A key step in breeding better crops that survive under stress has been to understand the changes in cellular, biochemical and molecular processes that occur in response to stress. Modern molecular techniques involve the identification and use of molecular markers that can accelerate the efficiency of breeding programs. However, the introgression of genomic portions (QTLs) involved in stress tolerance often carry undesirable agronomic characteristics from the donor parents. This is because of the lack of a precise knowledge of the key genes underlying the QTLs controlling the target traits. In this context, development of genetically engineered plants by the introduction and/or over-expression of selected genes seem to be a viable option to hasten the breeding of "improved" plants or in generating the precious pre-breeding materials. Intuitively, genetic engineering would be a faster way to introduce any target beneficial genes than through conventional or molecular breeding. Also, it would be the only option when genes of interest originate from cross barrier species, distant relatives, or from non-plant sources. Following these logical steps, various transgenic technologies have been used to improve stress tolerance in plants (Allen 1995).

Problems associated with engineering resistance through genetic transformation is that; there is a possibility that pests may evolve resistance to the transgenes used against them, rendering the transgenic plants susceptible to pest attack. In this context, gene pyramiding or simultaneous transfer of multiple genes may be helpful in conferring broad spectrum resistance against different races of pathogens or conferring resistance against different pests or pathogens or combination of stresses. Essentially gene pyramiding enable us to have simultaneous expression of more than one gene(s) associated with resistance in a transgenic plant (Shelton et al. 2002).

The rationale behind gene pyramiding originates from the age old philosophy of the use of insecticide mixtures to broaden the spectrum of insects controlled in one spray event. Gene pyramiding has been hailed as a lasting insect/disease resistance management strategy (Jackson et al. 2003; Shelton et al. 2002). However, a closer look at the strategy reveals that pyramiding was developed as a practical strategy to broaden the range of insect species that were not being adequately controlled by a single toxin as in the case of the single gene Bollgard® Bt cotton variety. It is also believed that gene pyramiding strategy can be successfully deployed against combination pest and diseases and in some cases it can be successfully implemented against combination of biotic and abiotic stresses.

7.2 Gene Pyramiding for Biotic Stress Tolerance

Plant pathogens and phytophagous pests are the major cause of agricultural losses in both developed and underdeveloped countries. Sustainable strategies for their management include breeding for resistance, application of chemical compounds as well

as cultural and biological control methods. Currently, these approaches have been complemented by plant biotechnology and genetic engineering for the development of genotypes more resistant to pests, phytopathogenic fungi and other agents of biotic stresses (Campbell et al. 2003; Ferry et al. 2006; Ranjekar et al. 2003). Biotechnology has allowed the development of novel strategies to obtain plants that are more resistant to pests, fungal pathogens and other agents of biotic stress.

7.2.1 Gene Pyramiding for Disease Resistance

Conventionally, genetic engineering means genetic transformation of the crop with a single gene associated with the trait of interest (Ronald 1997; Zhang et al. 1998). As disease resistance is a highly complex multigenic trait, generally single gene transformations result in insufficient and/or narrow spectrum disease resistance (Neuhans et al. 1991; Anand et al. 2003). In addition, there is always a possibility for reversal of resistance because of appearance of resistant strains of pathogens (Mew et al. 1992). Hence, genetic engineering of crop plants with (i) a combination of genes encoding/controlling interdependent or synergistic sub-components of disease-resistance to realize effective resistance against a particular disease and (ii) a combination of genes conferring resistance against different diseases to realize a wide-spectrum disease resistance would be more logical. A well planned genetic engineering strategy involving a well balanced expression of transgenes with different modes of action would ensure enhanced and durable resistance against different pathogens at a time. Availability of genes associated with disease resistance (Song et al. 1995) or genes involved in various defense mechanisms (Huang et al. 1994; Velazhahan et al. 1998; Takakura et al. 2000) and availability of efficient transformation tools and strategies (Christou 1997; Chen et al. 1998; Kim et al. 2003) has made pyramiding genes involved in disease resistance in crop plants much easier.

The success of multiple gene transformation or gene pyramiding is largely dependent upon the judicious selection of the genotypes/varieties to be transformed in a particular crop and the genes to be transformed together. Selection of suitable genes encoding proteins involved in pathogen recognition and/or subsequent activation of signaling pathways leading to activation of defense responses (Van Loon and Van Strien 1999) and genes exhibiting direct antimicrobial properties (Broglie et al. 1991; Yun et al. 1998) by their degradative interaction with pathogen cell wall and cell membrane can be expected to improve the scope for generating crop plants with specific or wide-spectrum resistance against pathogens.

Plants perceive signals (elicitors) from pathogens which initiates host-pathogen interactions. Recognition of pathogen's elicitors triggers multiple defense responses in hosts, including the accumulation of defense compounds called phytoalexins, pathogenesis-related (PR) proteins, evolution of reactive oxygen species and hypersensitive cell death. PR-proteins accumulate in plant cells along with other defense-related molecules during plant's hypersensitive response (HR) (Heath 2000) against

pathogen attack. PR-proteins are classified on the basis of their amino acid sequences, serological relationships and biochemical functions (Van Loon and Van Strien 1999) and several members of the PR-protein categories are well studied by various groups (Muthukrishnan et al. 2001). It has been demonstrated that constitutive and over-expression of PR-proteins in transgenic plants resulted in enhanced resistance against a variety of pathogens.

Different endochitinases (EC 3.2.1.14) belonging to PR-3, 4, 8, and 11 sub group of PR-proteins (Van Loon and Van Strien 1999) catalyze the hydrolysis of β-1,4 linkages between N-acetyl-glucosamine units of chitin, a major polysaccharide component of fungal cell-wall (Broglie et al. 1991). A few chitinase cDNAs have been isolated from rice (Huang et al. 1991; Takakura et al. 2000) and have been employed to transform rice cultivars to engineer sheath blight resistance (Lin et al. 1995; Datta et al. 2001; Kumar et al. 2003; Kalpana et al. 2006). Expression of genes encoding this protein in plants such as apple, broccoli, carrot, cucumber, peanut, sorghum, strawberry, tobacco, wheat, bent grass, rice and silver birch, provided varying levels of enhanced resistance to different fungal pathogens (Punja, 2006).

Thaumatin like Proteins (TLP) belong to group of PR proteins which causes lysis of fungal cell membranes leading to death of the pathogen by altering the cell membrane permeability (Yun et al. 1998). Examples of thaumatin like proteins include thaumatin (*Thaumatococcus daniellii)*, Osmotin (*Nicotiana tabacum*), Zeamatin (*Zea mays*) and permeatin (cereal). Several members of this protein family possess antifungal properties and they are categorized under PR-protein group 5 (PR5). The transgenic rice plants over-expressing a rice *tlp* transgene have been reported to exhibit an enhanced resistance against sheath blight pathogen (Datta et al. 1999; Krishnan Kalpana et al. 2006). A subgroup of TLPs has been shown to be β-1, 3-glucan binding proteins (Osmond et al. 2001).

Disease resistance is a complex trait controlled by several groups of genes. Hence, constitutive expression of a single PR-protein transgene like *tlp* could not be expected to confer sufficient level of disease resistance. The marginal and narrow spectrum disease resistance conferred by single PR-protein is one of the major reasons for the reported failures. However, the co-expression of more than one PR-protein genes such as chitinase (EC 3.2.1.14) and β-1, 3-glucanase (EC 3.2.1.39) was shown to be much more effective against several fungal diseases. Krishnan Kalpana et al. (2006) showed co-transformation of genes encoding rice chitinase and thaumatin-like protein (TLP) in an elite *indica* rice line and the transformants exhibited significantly higher level of sheath blight (ShB) resistance than the transgenic plants introduced with either chitinase or TLP alone. Some of the popular indica rice cultivars namely ADT38, ASD16, IR50 and PB1 were transformed with *tlp* and *chi11* (encoding a 35 kDa rice chitinase) in addition to transgenics generated by introducing *tlp* and *chi11* genes independently. Level of resistance against sheath blight and sheath rot diseases was compared between the *tlp* + *chi11* transformants with that of *tlp* transformants. Stable inheritance of the transgene expression was studied up to T_2 generation by western blotting analysis. The putative transformants introduced with *tlp* showed enhanced resistance against the sheath

blight pathogen, *Rhizoctonia solani* when compared to the non-transformed plants. The transgenic plants introduced with both *chi11 and tlp exhibited* enhanced resistance against *R. solani* than the ones that express *tlp* or *chi11* transgene alone.

In the similar fashion, Maruthasalam et al. (2007) reported pyramiding of three genes namely *chi11*, *tlp*, and *Xa21* into popular indica rice cultivars viz., Pusa basmati-1, ASD16, ADT38, IR72, and White Ponni to engineer effective and stable resistance against a fungal pathogen causing Sheath blight disease and a bacterial pathogen causing blight disease (BLB) in rice. Calli derived from these rice cultivars were co-transformed with these three genes through particle bombardment method. Molecular analyses of putative transgenic lines by PCR, Southern hybridization analysis and Western Blotting analysis revealed stable integration and expression of the transgenes in few independent transgenic lines. A transgenic Pusa basmati-1 line pyramided with *chi11*, *tlp*, and *Xa21* was identified and found to exhibit enhanced resistance against both sheath blight and bacterial blight.

Corrado et al. (2007) reported that a Chitinase A protein from *Autographa californica* Nuclear Polyhedrosis Virus (AcMNPV) is capable of conferring enhanced resistance against both fungal pathogens and herbivorous pests in tobacco. Transgenic tobacco plants showing over-expression of an active *Chi-A* protein showed reduced damages caused by fungal pathogens and lepidopteran larvae. This is the first report on characterization and expression of a single gene introduced into plants that increases resistance against herbivorous pests and fungal pathogens. Co-transformation of an insect resistant gene namely *gna (agglutinin) along with Xa21* gene has been demonstrated in rice to improve both pest and disease resistance in rice (Tang et al. 1999). In few instances, attempts have been made through combination of marker-assisted breeding (MAS) and genetic transformation strategies to develop rice lines resistant to blast and bacterial leaf blight (BLB) by pyramiding *Pi1 and Piz5* (major blast resistance genes) through marker assisted breeding and introducing *Xa21* gene through genetic transformation (Narayanan et al. 2004).

Datta et al. (2002) reported a novel strategy of pyramiding three genes involved in conferring pest and disease resistance in rice. They developed an IR 72 transgenic line introduced with Bt cry gene and another IR 72 line introduced with *Xa21 and RC7 chitinase* gene. These two independent lines were crossed through conventional breeding to develop a single elite rice line possessing both insect and disease resistance. Dong et al. (2007) reported engineering resistance in tall fescue against two major fungal diseases namely gray leaf spot caused by *Magnaporthe grisea*, and brown patch caused by *Rhizoctonia solani*. Three different genes (β-1, 3-glucanase from alfalfa (AGLU1), a truncated dermaseptin *SI* gene from the South American arboreal frog *Phyllomedusa sauvagei* and *Pi9* gene which is an *R* gene conferring resistance against grey leaf spot) from various sources were introduced into two elite cultivars, Coronado' and 'Matador' of tall fescue through *Agrobacterium*-mediated transformation and the transgenics were found to exhibit resistance against these two major diseases in tall fescue.

7.2.2 Gene Pyramiding for Insect Resistance

The use of gene transfer technology to introduce insect-resistance genes into crop plants provides an economical and environmentally sustainable alternative to the extensive use of chemicals for the control of insect pests. In USA, introduction of transgenic crops in major crops viz., canola, corn, cotton, papaya, squash and soybean resulted in increased production of additional 2.4 million tonnes of food and fibre and increased farm income by US$1.9 billion. In addition, introduction of these biotechnology-derived crops reduced the use of pesticides by 21,000 tonnes (Sankula et al. 2005). Transgenic crops expressing individual *Bacillus thuringiensis* (*Bt*)-endotoxin genes have been commercially available for several years, while other insecticidal genes, notably those encoding protease inhibitors and lectins are undergoing experimental trials (Gatehouse et al. 1997; Graham et al. 1997). A hybrid commercial *Bt* rice, Shan You 63, has now been field-evaluated in China on a large scale and has shown resistance to two insect pests, leaffolder and yellow stem borer (Tu et al. 2000). Transgenic IR72 with fusion *Bt* also showed very high protection against four lepidopteran insects under Chinese field conditions (Ye et al. 2001).

One problem that could arise due to extensive use of insect-resistant transgenic plants is the evolution of resistance in different pest populations. A number of strategies (Alstad and Ardow 1995; McGaughey et al. 1998) have been proposed to delay or prevent such outcome. Simultaneous deployment of two or more plant insect resistance traits may be an effective way to inhibit pest adaptation, even with smaller and more economically acceptable refuge sizes. Mixtures of resistance genes can be achieved by inter-planting two cultivars, each with a different genetic basis for resistance, or by introducing two different resistance genes into the same cultivar (gene pyramiding). A viable strategy to overcome such undesirable effect involves the simultaneous introduction of several resistance genes with different modes of action against the same pest into the same plant (Ruud et al. 1999) or introduction of different genes conferring resistance against different pests into the same plant which is expected to offer broad spectrum resistance against different biotypes of same insect pest or different insect pests of same crop.

Transgenic plants that express *Bt cry* endotoxins or protease inhibitors were found to exhibit resistance against insects that feed by chewing or biting (Lepidopterans and Coleopterans). There is little evidence that such plants confer resistance against sap sucking pests (Homoptera) like brown plant hopper (BPH) and green leaf hopper (GLH). Certain plant lectins including snowdrop lectin (*Galanthus nivalis* agglutinin; *gna*) were found to possess toxicity against these homopteran insects (Powell et al. 1995) as well as lepidopteran and coleopteran larvae (Gatehouse et al. 1995). Transgenic rice plants capable of producing *Bt cry* proteins *cry1Ab* and *cry1Ac* were found to confer greater degree of resistance against their target pests namely striped stemborers and yellow stem borers (Cheng et al. 1998; Datta et al. 1998). Maqbool et al. (1998) were able to demonstrate the improved level of insect resistance of Basmati 370 and M7 rice lines engineered

with *cry2A* gene and the insect bioassays demonstrated the effectiveness of Cry2A protein against yellow stem borer and rice leaf folder in rice.

The simultaneous introduction of three genes expressing insecticidal proteins, *Cry1Ac, Cry2A* and *gna* into indica rice to control three major pests, rice leaf folder (*Cnaphalocrocis medinalis*), yellow stemborer (*Scirpophaga incertulas*) and the brown planthopper (*Nilaparvata lugens*), has also been reported (Maqbool et al. 2001). The *cry* genes target the leaf folder and the stem borer, and the *gna* gene targets the planthopper. Triple transgenic plants were found to exhibit more resistance compared with their binary counterparts. Comparison of three different transgenic Bt cotton populations containing either the single Cry1Ac or Cry2Ab gene, or both genes, for fruit penetration and damage by cotton bollworm larvae (Jackson et al. 2004). These are few examples where transgene pyramiding was used in a crop plant to create durable resistance against multiple insect pests with different feeding modes.

Computer simulations predict that gene pyramiding is much more effective in the delay of the evolution of a resistant insect population even with a smaller refuge size (Roush 1997). Some experimental data on the simultaneous deployment of multiple insect resistance genes (Greenplate et al. 2000; Stewart et al. 2001) demonstrated that cotton producing two *Bt* gene products had greater insecticidal activity than cotton that carried only the *cry1Ac* gene. Zhao et al. (1999) evaluated transgenic tobacco plants expressing Bt *cry1Ac protein* and a gene encoding a cowpea trypsin inhibitor (*CPT1*) and demonstrated that use of transgenics carrying both the genes resulted in delayed development of resistance in *Helicoverpa armigera* in comparison to plants carrying only the *cry1Ac* gene.

Cao et al. (2002) demonstrated a novel way of producing insect resistant plants exhibiting higher level of resistance against diamondback moths in broccoli by combining both genetic transformation and conventional breeding methods. They generated independent transgenic broccoli plants harboring Bt *cry1Ac* and *cry1C* genes in broccoli and these genes were pyramided into a single broccoli line plant by effecting hybridization between them. Broccoli lines carrying both *cry1Ac* and *cry1C Bt* genes were selected by screening with kanamycin and hygromycin and pyramiding was confirmed by PCR analysis. Stable expression of *Cry1Ac* and *Cry1C* proteins was confirmed in the pyramided line which was found to cause rapid and complete mortality of diamondback moth larvae. In the similar fashion, Cao et al. (2008) adopted the strategy of sequential transformation to pyramid two different *Bt* genes (*cry1Ac* and *cry1C*) in Indian vegetable mustard (*Brassica juncea* L.) to control diamondback moth larvae. First *cry1C* was introduced into Brassica through *Agrobacterium*-mediated transformation. A *cry1C transgenic* line was then transformed with the *cry1Ac* gene to produce pyramided *cry1Ac* + *cry1C* plants. Six plants were found to contain both *cry1Ac and cry1C* genes which was confirmed by PCR and Southern hybridization analysis and the stable expression of the protein were confirmed by ELISA screening. Insect bioassays indicated that transgenic plants pyramided with both the genes *cry1Ac and cry1C* showed toxicity against all three types of diamondback moth larvae and lepidopteran insect pests.

Chitinolytic enzymes are interesting candidates for plant protection as they are active against many noxious pests and fungi and because of their non-toxicity against non target organisms that lack chitin (i.e. vertebrates and plants). Chitin is a key component of the cell wall of the majority of fungal phytopathogens and it is present in the cuticle and the peritrophic membrane (PM) lining the midgut of several insect species. Chitinolytic enzymes act on the PM of pests feeding on plants that express these proteins and impair the gut physiology (Kramer and Muthukrishnan 1997). One such chitinolytic enzyme called chitinases proved to be effective in enhancing plant resistance in conjunction with other known insecticidal proteins or toxins (Schuler et al. 1988; Wang et al. 2005).

Transgenic plants expressing protease inhibitors have to date shown marginal effectiveness against insect pests. Protease Inhibitors have the potential to be effective insecticidal proteins if insect adaptation to them can be overcome. The use of novel inhibitors, such as the barley trypsin inhibitor (Alfonso-Rubi et al. 2003), equistatin from sea anemone (Gruden et al. 1998), other cystatins (Martinez et al. 2005) or synthetic constructs containing multiple inhibitors (Outchkourov et al. 2004) or inhibitors and lectins might also prove useful. Second generation insect-resistant transgenic plants with increased potential for durable resistance might result from the deployment of plants expressing multiple insecticidal novel proteins such as the Vip (vegetative insecticidal proteins) produced by *Bacillus thuringiensis* during its vegetative growth. These have insecticidal activity towards a wider spectrum of insect pests and transgenic cotton expressing such a Vip is under commercial cultivation in the USA. Considerable progress has also been made in the identification of several toxin genes from bacterias namely Photorhabdus and Xenorhabdus which are symbionts of entomopathogenic nematodes (Williamson and Kaya 2003). These genes encode large insecticidal toxin complexes with little homology to other known toxins. These toxins cause septicaemia in the insect, the insect is killed and its tissues are used as nutrients by the nematode. Arabidopsis plants expressing toxin A gene from *Photorhabdus luminescens* showed strong insecticidal activity against one lepidopteran and moderate activity against a coleopteran pest (Liu et al. 2003). These new candidate genes can be used for genetic transformation studies to pyramid multiple resistances against various insect pests in crop plants.

7.3 Abiotic Stress Tolerance in Plants

It has been estimated that abiotic stresses such as drought, salinity, extreme temperatures, submergence and oxidative stress reduce crop yield worldwide by more than 50% (Bray et al. 2000). Drought is the major abiotic stress affecting 26% of the arable area followed by mineral toxicities /deficiencies including salinity. The fraction of world's arable land subjected to various kinds of abiotic stresses including 26% drought, 20% mineral and 15% freezing. These stresses are often interconnected

cause similar cellular damage and activate similar cell-signaling pathways (Shinozaki and Yamaguchi-Shinozaki 2000; Zhu 2001; Zhu 2002).

Drought resistance can be introgressed into any desired genotypes by suitable breeding procedures like backcross breeding method and using donors for various genes in tandem or by opting for convergent crossing incorporating all the desired genes into the commercial recipient parent. But selection procedure in the large number of segregating population is tedious because of the complexity in phenotyping. Targeted introgression and pyramiding of multiple genes from diverse parents requires lot of manpower and time for thorough and accurate phenotyping. Salinity affects virtually all aspects of a plants physiology and the survival of the plant under saline conditions involves multiple adaptations. Despite considerable efforts through both international and national breeding programs, progress in enhancing tolerance to salinity has been slow with only few new cultivars released because of the complexity in mechanisms controlling this trait.

Low temperature stress leads to formation of ice crystals within the intercellular spaces which results in severe osmotic and mechanical injury to the plant. In addition, freezing can induce the production of reactive oxygen species (ROS) which damages the membrane components leading to denaturation of cellular proteins (Thomashow 1999). The similarities between the consequences of drought and cold stress are clearly evident from the plant responses viz., activation of antioxidant defense mechanisms, production of heat shock proteins (HSPs), synthesis of high concentrations of intracellular praline and accumulation of sugars.

7.3.1 Genetic Engineering Strategies to Improve Abiotic Stress Tolerance in Plants

Abiotic stress factors such as drought, salinity and extremes of temperature have long been known as major limitations to crop productivity (Boyer 1982). One approach to improve stress tolerance in crops would be to transfer the abiotic stress tolerance related adaptive traits from the tolerant organism. However, this process has limited success using conventional means (Yeo and Flowers 1989) mainly due to poor understanding of genetic and physiological basis of abiotic stress tolerance in plants and partly due to transfer of unwanted genes during conventional crossing. In this context, genetic transformation technology enables us to achieve transfer of genes controlling the target traits involved in abiotic stress tolerance in a precise and predictable manner.

When plants are subjected to abiotic stress, a number of genes are turned on, resulting in increased levels of several metabolites and proteins involved in protective functions. These genes are the main targets for engineering abiotic stress tolerance in crop plants. Stress-induced gene expression can be broadly categorized into three groups: (1) genes encoding proteins with known enzymatic or structural functions, (2) proteins with as yet unknown functions and (3) regulatory proteins. Initial attempts were made to develop transgenics for abiotic stress tolerance involving

“single action genes” i.e., genes responsible for modification of a single metabolite that would confer increased tolerance to salt or drought stress. Metabolic traits especially pathways with few enzymes are better characterized genetically and more amenable to such manipulations than structural and developmental traits. Stress-induced proteins with known functions such as water channel proteins, key enzymes for osmolyte (proline, betaine, sugars such as trehalose, and polyamines) biosynthesis, detoxification enzymes and transport proteins were used for plant genetic transformation experiments to improve abiotic stress tolerance in plants (Bajaj et al. 1999; Apse and Blumwald 2002). Few examples for generation of such abiotic stress tolerant transgenic plants are listed in Table 7.1.

However, the above-said approach has overlooked the fact that abiotic stress tolerance is likely to involve many genes at a time and this single-gene manipulation of abiotic stress tolerance is unlikely to be sustainable. Therefore, a second “wave” of genetic transformation was made to transform plants with the third category of stress-induced genes namely, regulatory proteins and signaling proteins. Through these proteins, many genes involved in stress response can be simultaneously regulated by a single gene encoding stress inducible transcription factor (Kasuga et al. 1999), thus offering possibility of enhancing tolerance against multiple stresses including drought, salinity, and freezing.

Recently, genes involved in stress signal sensing are targeted for engineering abiotic stress tolerance in plants (Winicov 1998; Shinozaki and Yamaguchi-Shinozaki 1999). By regulating the expression of a regulatory gene that could induce a number of down stream genes involved in stress-tolerance, transgenic plants with a stress-tolerant phenotype could be achieved. For example, multiple stress resistance in *A. thaliana* was pyramided by engineering stress-inducible expression of the transcription factor DREB1A which resulted in improved drought, salt and freezing tolerance (Kasuga et al. 1999). In another attempt, Ito et al. (2006) analyzed OsDREB1 transgenic rice plants and confirmed its improved tolerance to drought, salt and low temperatures and identified large portion of stress-inducible genes. These results confirm that DREB1/CBF cold-responsive pathway is conserved in rice and Arabidopsis.

Next to DREB family transcription factors, ABA-responsive element binding factor (ABF) members belonging to bZIP transcription factor family seems to be promising candidates to improve abiotic stress tolerance in plants. ABFs show distinct roles in sugar, ABA and stress responses (Uno et al. 2000). To reveal the involvement of ABFs in stress tolerance, Kang et al. (2002) generated ABF3 and ABF4 transgenic Arabidopsis lines by over-expressing them constitutively. ABF3 and ABF4 over-expression mutant lines exhibited an altered transpiration rate in response to water deficit conditions, eventually all transgenic plants survived a 12-day drought treatment and set seed in contrast to wild type plant with 33% survival rate. Both ABF3 and ABF4 over-expression lines showed induction of ABA signaling ABI1, ABI2 phosphatase and other stress-responsive genes including desiccation-related LEA genes via ABA dependent pathway. In contrast, both ABF3 and ABF4 transgenic lines were hypersensitive to salinity treatments during germination. In another study made by Kim et al. (2004), in contrast to ABF3 and ABF4,

Table 7.1 Abiotic stress tolerant transgenic plants produce by engineering metabolic pathways/metabolite accumulation

Gene/function	Source of the gene	Performance of transgenic plant	References
GS2/glutamine synthase	Rice	Tolerant to salt and possibly chilling stresses	Hoshida et al. (2000)
OsCDPK7/calcium-dependent protein kinase	Rice	Tolerant to salt and drought stresses	Saijo et al. (2000)
OsMAPK5/MAP kinase	Rice	Inverse modulation of biotic and abiotic stress tolearance by MAP kinase	Xiong and Yang (2003)
YK1/rice homolog of maize HC toxin reductase	Rice	Tolerant to multiple abiotic stresses including salt and submergence	Uchimiya et al. (2002)
OsNHX1, AgNHX1/	Rice,	Tolerance to salt stress	Ohta et al. (2002);
Na^+/H^+ antiporter *HVPIP2*; *1*/aquaporins	*Atriplex gmelini*, Rice Barley	Differential response to salt-stress	Fukuda et al. (2004);
Adc, *Samdc*/polyamine biosynthesis	Datura/oat/ tritordeum	Salt/drought tolerant	Capell et al. (2004)
HVA1/LEA proteins	Barley, wheat	Tolerant to salt and dehydration stress	Babu et al. (2004)
codA/glycine betaine	*Arthrobacterglobiformis*	Tolerant to salt and/or drought stress	Sawahel (2003)
OtsA, *OtsB*/trehalose biosynthesis	*E. coli*	salt, drought and cold	Garg et al. (2002);
P5cs/proline biosynthesis	Mothbean	Stress induced transgene expression	Su and Wu (2004);
pdc1, *adc*/pyruvate decarboxylase/alcohol dehydrogenase	Rice	Submergence tolerant	Quimio et al. (2000);
AGPAT, *SGPAT*/fatty acids biosynthesis	*Arabidopsis*, spinach	Improved photosynthesis and growth at low temperatures	Ariizumi et al. (2002)
Cat/catalase	Wheat	Cold tolerant	Matsumura et al. (2002)
GST/ζGlutathione S-transferase	Rice	Enhanced germination and growth under low temperature	Takesawa et al. (2002)
Spl7/heat-stress transcription factor	Rice	Transformation of rice *spl7* mutant with *Spl7* show tolerance to heat stress	Yamanouchi et al. (2002)
hsp101, *hsp17.7*/heat-shock proteins	Rice	Tolerant to high temperatures	Katiyar-Aggarwal et al. (2003);

adc, arginine decaboxylase; *codA*, choline oxidase; hsp, heat-shock protein; GPAT, glycerol-3-phosphate acyl transferase; LEA, late embryogenesis-abundant protein; MAPK, mitogen activated protein kinase; *Naat*, nicotianamine aminotransferase; *OtsA* and *OtsB*, *E. coli* trehalose biosynthesis genes encoding for trehalose phosphate synthase (TPS) and trehalose phosphate phosphatase (TPP); pdc, puruvate decarboxylase; P5CS, Δ-pyrroline-5-carboxylste synthase; *Spl*, spotted leaf.

the two-week old ABF2 transgenic plants exhibited higher salt tolerance (Kim et al. 2004). These studies are providing enough evidences for the involvement of the ABF family members in different abiotic stress tolerance pathways which seems to be good candidates for gene pyramiding strategy of improving abiotic stress tolerance in plants.

Stress perception and signaling pathways are critical components of adaptive response that is vital for the plant species to survive under extreme environmental constraints (drought, salt and extreme temperatures). Osmotic stress and associated oxidative stress are a common consequence of such a stress exposure and share one or more intermediates/components or outputs as part of their signaling (Viswanathan and Zhu 2004). It is also highly desirable to have stress signaling sensors that can transduce the signal to the target cells. Although signal perception and transduction pathway genes are attractive targets for genetic engineering, a detailed knowledge of cascades of signal perception and transduction activated under abiotic stress response is lacking. Few reports are providing us clues to target these candidates to improve abiotic stress tolerance in plants. A receptor-like protein kinase (NtC7) induced under abiotic stress response in tobacco has been shown to confer osmotic stress tolerance in over-expressing transgenic tobacco plants (Tamura et al. 2003). Mitogen-activated protein kinase (MAPK) cascades seem to be the convergent points for crosstalk and it is shown that constitutive expression of MAPKKK/Nicotiana protein kinase 1 (MAPKKK/NPK1) in maize activated an oxidative signal cascade and the transgenics showed tolerance to cold, heat, salinity and also higher photosynthetic rates (Shou et al. 2004). Pardo et al. (1998) succeeded in developing a stress-tolerant transgenic plants by over-expressing calcineurin, a protein phosphatase known to be involved in salt-stress signal transduction in yeast. Interestingly, a calcium-dependent protein kinase (CDPK) has been identified as an important component of osmotic signaling pathways. Rice CDPK7 gene has been over-expressed and shown to be a positive regulator in triggering stress-responsive genes in response to salt/drought (Saijo et al. 2000). As indicated above, until now we have only fragmentary knowledge about the abiotic stress signaling pathways, which could be dissected in a systematic manner by forward and reverse genetics approach.

Further, genetic engineering allows controlling of timing, tissue-specificity and expression level of the introduced genes for their optimal function. This is an important consideration if the action of a given gene or transcription factor is desired only at a specific time, in a specific organ or under specific conditions of stress. The basic findings on stress inducible promoters have led to a major shift in engineering abiotic stress-tolerant crops during recent years (Katiyar et al. 1999).

7.4 Marker Assisted Gene Pyramiding

When transferring a specific desirable gene conferring resistance against pest/disease/abiotic stress from a wild plant into a crop plant via wide crosses, the simultaneous transfer of undesired genes from the wild plant is often a problem.

In repeated crosses to the cultivated type, the percentage of wild type genes in each generation would be 50% in the first generation, 25% in the second, 12.5% in the third, 6.25% in the fourth, etc. Even after many generations, wild-type genes located close to the desired gene on the chromosome may still be present (i.e., are closely *linked* to the desired gene).

Recently there have been advances in the mapping of genes involved in the variation of quantitative traits, through quantitative trait loci (QTL) mapping experiments and analysis of genomic data. Such studies on complex traits have led to the identification of a great number of genetic factors responsible for the heritable variation of these traits. Exploitation of molecular markers tightly linked to these genetic factors can be useful in assessing the genotype of corresponding individuals carrying favorable alleles at these loci which will provide genetic material for the development of new improved varieties. This led to the development of a novel strategy called "marker assisted selection" (MAS).

Marker-assisted (or *molecular-assisted*) *breeding* provides a dramatic improvement in the efficiency with which breeders can select plants with desirable combinations of genes. A *marker* is a "genetic tag" that identifies a particular location within a plant's DNA sequences. Markers can be used in transferring a single gene into a new cultivar or in testing plants for the inheritance of many genes at once. Markers can be based upon either DNA or proteins. Marker-assisted breeding is helpful to identify plants that have inherited the desirable gene together with as few of the undesired genes as possible. Traditional methods require sorting through the entire haystack to find the needle, while markers provide a "metal detector" to easily find the specific part of the haystack where the needle is located. As the number of traits (genes) that are desired increases, the number of plants that must be screened to identify plants that have the rare superior combinations of genes increases even more rapidly. By providing quick and efficient tests for many different genes, DNA markers are being widely used in breeding crop varieties having optimal combinations of desirable genes. DNA markers have been used for transferring pest resistance genes to cultivated varieties, assisting selection of complex multigene traits (such as flavor), aiding evaluation of regionally and seasonally optimized varieties, allowing genetic purity testing, and enabling proprietary variety protection and patent enforcement.

Most theoretical work on the application of marker-QTL associations in selection has focused on using markers to estimate an individual's breeding value more reliably than when using its phenotype. Advent of marker assisted selection strategy has made it possible to transfer locations of a series of genes of interest (hereafter referred to as target genes) which is otherwise termed as "gene pyramiding". Using marker assisted pyramiding strategy now it is possible to introgress two or three QTLs/genomic regions controlling various physiological/biochemical traits associated with stress resistance in plants. MAS help in efficient selection of target loci, minimizing linkage drag and rapid recovery of recurrent genome. A MAS strategy can also be used to introgress more than two traits efficiently into a common recurrent parent. It is reported that simultaneous transfer of upto 5 QTLs is possible through marker assisted gene pyramiding strategy (Fig. 7.1).

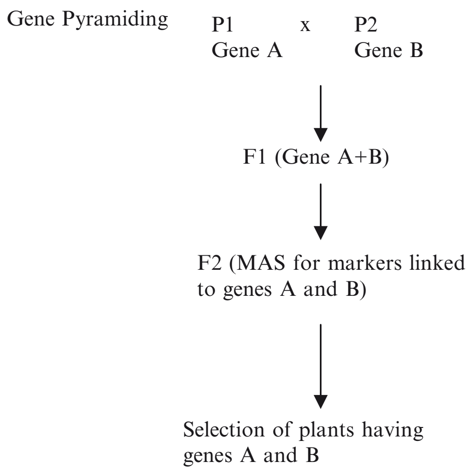


Fig. 7.1 A simple scheme for marker assisted pyramiding of two different traits

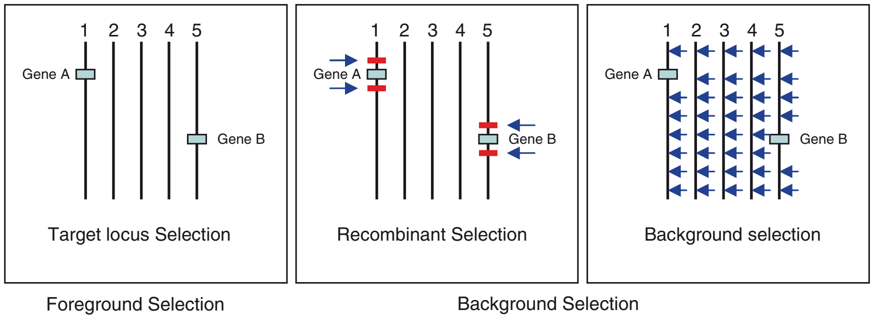


Fig. 7.2 Schematic diagram showing whole genome selection process in introgressing a target loci into a desired background

Implementation of marker assisted selection strategy in breeding programs will help in selecting for both target loci (from the donor parent) and background genome (recurrent parent). By this way it is possible to achieve maximum recovery of recurrent genotype (Fig. 7.2). The basis of a marker-assisted backcrossing (MAB) strategy is to transfer a specific allele at the target locus from a donor line to a recipient line while selecting against donor introgressions across the rest of the genome. The use of molecular markers, which permit the genetic dissection of the progeny at each generation, increases the speed of the selection process, thus increasing genetic gain per unit time (Tanksley et al. 1989). The effectiveness of marker assisted breeding depends on the availability of closely linked markers and/or flanking markers for the target locus, the size of the population, the number of backcrosses and the position and number of markers for background selection (Frisch and Melchinger 2005).

If markers tightly linked to different traits are available, it is possible to monitor introgression of many traits (upto six traits) at a time. A schematic representation for marker assisted pyramiding of six different traits (root and stem characters) is given in the Fig. 7.3 shown below (Servin et al. 2004).

Sharma et al. (2004) demonstrated the efficiency of MAS in pyramiding two Brown planthopper (BPH) resistance genes Bphl and Bph2 independently derived from two indica resistance lines into a single line through the recombinant selection. Markers linked tightly to the Bph locus were used for selecting the recombinants and the pyramided lines were found to exhibit higher level of resistance. Hittalmani et al. (2000) adopted the strategy of marker assisted QTL pyramiding to improve blast resistance in rice. They identified three major genes (Pil, Piz-5 and Pita) for blast resistance in rice located on chromosomes 11, 6 and 12 respectively. They fine mapped the corresponding regions which resulted in the identification of closely linked RFLP markers. The three genes were pyramided using RFLP markers and the plants carrying the two- and three-gene combinations were tested for resistance to leaf blast and the pyramided lines were found to have enhanced resistance than when it is present alone. In the similar manner Barloy et al. (2007) pyramided two genes from *Aegilops variabilis* conferring resistance against cereal cyst nematode into cultivated wheat. First, RAPD marker OpY16-$_{1065,}$ linked to CCN resistance was converted into a SCAR marker and used to introgress the regions into wheat. CCN bioassays showed that the level of resistance of the pyramided line

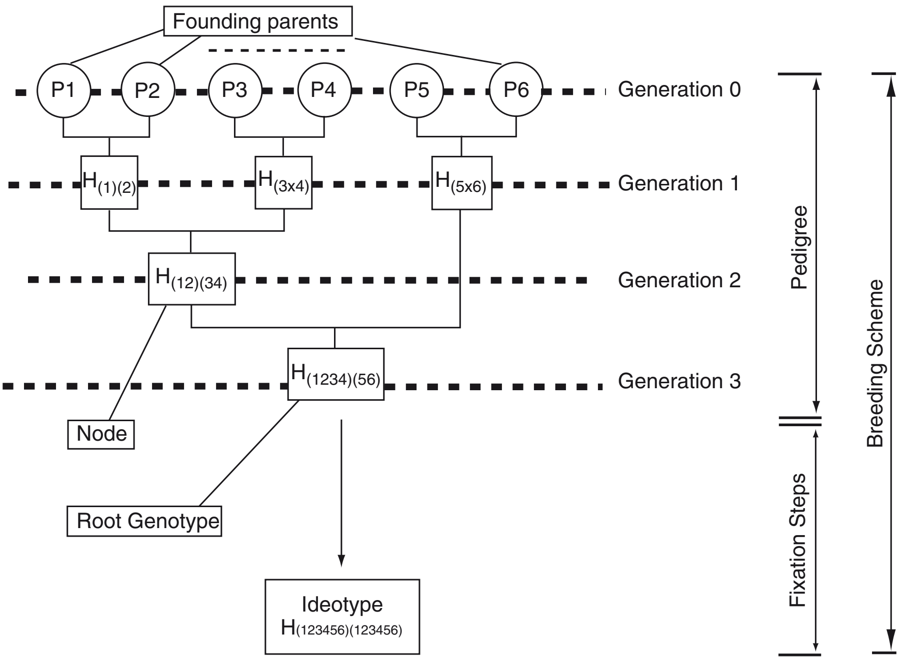


Fig. 7.3 Scheme of pyramiding multiple traits from diverse parents

was significantly higher than that of single introgression lines but lower than that of *Ae. variabilis*. This study thus illustrates the utilization of molecular markers in breeding for host resistance. Raman et al. (1999) reported the 'pyramided scald disease resistant lines' in barley significantly outyielded the susceptible cultivar, Clipper. They used DNA markers linked to scald resistance in AB35, AB200 and AB208-lines derived from *H. spontaneum*.

Werner et al. (2002) adopted a novel strategy of marker assisted pyramiding to introgress three different genes conferring resistance against Yellow Mosaic Virus in barley. For pyramiding rym4, rym9 and rym11 different strategies were employed exclusively based on doubled haploid lines (DHs) as homozygous recessive genotypes are more frequent in DHs than in segregating F_2-populations. In the DH lines produced from the F_1 of single crosses (*rym4* × *rym9*, *rym9* × *rym11*), plants that are homozygous recessive at both loci (25%) were identified by molecular markers. Respective genotypes were again crossed and F_1 derived DHs were screened by molecular markers. This strategy was found to be efficient in pyramiding resistance from multiple sources in a relatively short period of time. Maroof et al. (2008) developed a soybean line exhibiting resistance against various strains of soybean mosaic virus (SMV) through marker assisted gene pyramiding strategy. Simple sequence repeat markers were used to create resistant isogenic line in the background of susceptible cultivar Essex by pyramiding *Rsv1, Rsv3*, and *Rsv4* genes. They successfully demonstrated that MAS aided pyramiding is a straightforward method in generating isogenic lines with two- or three-genes to engineer high levels of resistance to SMV.

7.5 Conclusion

Gene pyramiding through genetic transformation can complement and accelerate breeding programs by introducing resistance genes from diverse sources, such as other plants, bacteria and even animals. It has several advantages: (1) Engineering of two or more genes for improving abiotic stress tolerance into a single cultivar is made possible in a gene-pyramiding breeding program. (2) The pyramided line can make use of any one of the introduced genes to escape or tolerate or resist the stress even if other genes do not work in favour. (3) Pyramiding approach involves less time period to develop multiple stress resistance in a crop when compared to conventional breeding or marker assisted QTL pyramiding. (4) The pyramided line serves as a pre-breeding material for all the genes and it can be used in any of the ensuing breeding schemes. (5) In the context of genetic diversity and gene pools, these pyramided lines act as a miraculous tool for conserving the rarest superior alleles. Apart from all these advantages, gene pyramiding strategy of improving stress resistance has few advantages viz. (1) Silencing of the gene(s) incorporated; (2) differential expression of the introgressed gene(s); (3) preservation of the genes corresponding to yield is a not a cakewalk; (4) it is necessary to develop efficient transformation and regeneration protocols which are strongly genotype dependent

in most crop plants and (5) screening for moisture stress, cold stress and salinity under green house conditions might not go in hand with the real microclimatic and macroclimatic environment in the field. From the available theoretical and practical knowledge, it is clear that gene pyramiding strategy is useful in broadening the genetic basis of resistance and this gene pyramiding can be successfully deployed in an integrated approach along with all the other management strategies.

References

Alfonso-Rubi J et al. (2003) Transgenic expression of trypsin inhibitor Cme from barley in *indica* and *japonica* rice, confers resistance to the rice weevil *Sitophilus oryzae*. Transgen Res 12:23–31

Allen RD (1995) Dissection of oxidative stress tolerance using transgenic plants. Plant Physiol 107(4):1049–1054

Alstad DN and Ardow DA (1995) Managing the Evolution of Insect Resistance to Transgenic Plants. Science 268, 1894–1896

Anand A, Zhou T, Trick HN, Gill BS, Bockus WW, Muthukrishnan S (2003) Greenhouse and fieldtesting of transgenic wheat plants stably expressing genes for thaumatin-like protein, chitinase and glucanase against *Fusarium graminearum*, J Exp Bot 54:1101–1111

Apse MP, Blumwald E (2002) Engineering salt tolerance in plants. Curr Opin Biotechnol 13:146–150

Ariizumi T, Kishitani S, Inatsugi R, Nishida I, Murata N, Toriyama K (2002) An increase in unsaturation of fatty acids in phosphatidylglycerol from leaves improves the rates of photosynthesis and growth at low temperatures in transgenic rice seedlings. Plant Cell Physiol 43(7):751–758

Anand A, Zhou T, Trick HN, Gill BS, Bockus WW, Muthukrishnan S (2003) Greenhouse and field testing of transgenic wheat plants stably expressing genes for thaumatin-like protein, chitinase and glucanase against Fusarium graminearum, J Exp Bot 54:1101–1111

Babu RC, Zhang J, Blum A, David Ho TH, Wu R and Nguyen HT (2004) *HVA1*, a LEA gene from barley confers dehydration tolerance in transgenic rice (*Oryza sativa* L.) via cell membrane protection. Plant Science, 166 (4):855–862

Bajaj S, Tragolli J, Liu L-F, Ho T-HD, Ray W (1999) Transgenic approaches to increase dehydration-stress tolerance in plants. Mol Breed 5:493–503

Barloy D, Lemoine J, Abelard P, Tanguy AM, Rivoal R, Jahier J (2007) Marker-assisted pyramiding of two cereal cyst nematode resistance genes from *Aegilops variabilis* in wheat. Mol breed 20(1):31–40

Boyer JS (1982) Plant productivity and environment. Science 218:443–448

Bray EA et al. (2000) Responses to abiotic stresses. In: Gruissem W et al. (eds) Biochemistry and molecular biology of plants. American Society of Plant Physiologists Rockville, MD, USA. pp 1158–1249

Broglie K, Chet I, Holliday M, Cressman R, Biddle PH, Knowlton S, Mauvais CJ, Broglie R (1991) Transgenic plants with enhanced resistance to the fungal pathogen *Rhizoctonia solani*. Science 254:1194–1197

Campbell MM, Brunner AM, Jones HM, Strauss SH (2003) Forestry's fertile crescent: the application of biotechnology to forest trees. Plant Biotech J 1:141–154

Cao J, Zhao JZ, Tang JD, Shelton AM and Earle ED (2002) Broccoli plants with pyramided cry1Ac and cry1C Bt genes control diamondback moths resistant to Cry1A and Cry1Ac proteins. Theor Appl Genet., 105:258–264

Cao, J, Shelton A.M and Earle ED (2008) Sequential transformation to pyramid two Bt genes in vegetable Indian mustard (*Brassica juncea* L.) and its potential for control of diamondback moth larvae. Plant Cell Rep, 27:479–487

Capell T, Bassie L, Christou P (2004) Modulation of the polyamine biosynthetic pathway in transgenic rice confers tolerance to drought stress. PNAS 101(26):9909–9914

Chandrababu R, Zhang J, Blum A, David Ho, WU R, Nguyen HT (2004) HVA1, a LEA gene from barley confers dehydration tolerance in transgenic rice (*Oryza sativa* L.) via cell membrane protection. Plant Sci 166:855–862

Chen H-k, Li B-q et al. (1998) Transformation of sulfur during pyrolysis and hydropyrolysis of coal. Fuel 77(6):487–493

Cheng X, Sardana R, Kaplan H, and Altosaar I (1998) Agrobacterium-transformed rice plants expressing synthetic cryIA(b) and cryIA(c) genes are highly toxic to striped stem borer and yellow stem borer. Proc Natl Acad Sci USA 95:2767–2772

Christou P (1997) Biotechnology applied to grain legumes. Field Crop Res 53:83–97

Corrado G, Arciello S, Fanti P, Fiandra L, Garonna A, Digilio MC, Lorito M, Giordana B, Pennacchio F, Rao R (2007) The Chitinase A from the baculovirus AcMNPV enhances resistance to both fungi and herbivorous pests in tobacco. Transgenic Res 17(4):557–571

Datta K, Tu J, Oliva N, Ona II, Velazhahan R, Mew TW, Muthukrishnan S, Datta SK (2001) Enhanced resistance to sheath blight by constitutive expression of infection-related rice chitinase in transgenic elite indica rice cultivars. Plant SciFeb 5; 160(3):405–414

Datta K, Vasquez A, Khush GS, Datta SK (1999) Development of transgenic New Plant Type rice for stem borer resistance. Rice Genet Newsl 16:143–144

Datta K, Vasquez A, Tu J, Torrizo L, Alam MF, Oliva N, Abrigo E, Khush GS, Datta SK (1998) Constitutive and tissue–specific differential expression of CryIA(b) gene in transgenic rice plants conferring enhanced resistance to insect pests. Theor Appl Genet 97:20–30

Datta SK, Baisakh N, Balachandran S, Torrizo LB, Chandel G, Arboleda M, Abrigo EM, Oliva NP, Tu J, Datta K (2002) Development of transgenic Bt rice and its deployment in Asia. In: Akhrust AJ et al. (ed) Biotechnology of *BacilluS thuringiensis* and its environmental impact. CSIRO, Canberra, pp 151–153

Dong S, Tredway LP, Shew HD, Wang GL, Sivamani E, and Qu R (2007) Resistance of transgenic tall fescue to two major fungal diseases. Plant Sci., 173(5):501–509

Ferry N et al. (2006) Transgenic plants for insect pest control. A forward looking scientific perspective. Transgenic Res 15:13–19

Fukuda A, Nakamur A, Tagiri A, Tanaka H, Miyao A, Hirochika H, Tanaka Y(2004) Function, intracellular localization and the importance in salt tolerance of a vacuolar Na^+/H^+ antiporter from rice. Plant Cell Physiol 45:146–59

Gatehouse AMR, Davidson G, Newell CA, Merryweather A, Hamilton WDO, Burgess EJ, Gilbert RJC, and Gatehouse JA (1997) Transgenic potato plants with enhanced resistance to the tomato moth, *Lacanobia oleracea* growth room trials. Mol. Breeding, 3:49–63

Gatehouse AMR, Powell KS, Van Damme EJM and Gatehouse JA (1995) Insecticidal properties of plant lectins. In Lectins - Biomedical Perspectives (Pusztai A and Bardocz S, eds.). London, UK: Taylor and Francis.

Garg AK, Kim JK, Owens TG, Choi YD, Kochian LV, Wu RJ (2002) Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses. Proc Natl Acad Sci 25:15898–15903

Graham J, Gordon SC, McNicol RJ (1997) The effect of the CpTi gene in strawberry against attack by vine weevil (Otiorhynchus sulcatus F. Coleoptera). Ann. Appl. Biol. 131:133–139

Greenplate JT, Penn SR, Shappley Z, Oppenhuizen M, Mann J, Reich B, Osborn J (2000) Bollgard II efficacy: quantification of total lepidopteran activity in a two-gene product. Proc Beltwide Cotton Conf 2:1041–1043

Gruden K, Strukelj B, Popovic T, Lenarcic B, Bevec T, Brzin J, Kregar I, Herzog-Velikonja J, Stiekema WJ, Bosch D, Jongsma MA. (1998) The cysteine protease activity of Colorado potato beetle (Leptinotarsa decemlineata Say) guts, which is insensitive to potato protease inhibitors, is inhibited by thyroglobulin type-1 domain inhibitors. Insect Biochem Mol Biol. 28(8):549–60

Heath MC (2000) Hypersensitive response-related death. Plant Mol Biol 44:321–334

Hittalmani S, Parco A, Mew TV, Zeigler RS, Huane N (2000) Fine mapping and DNA marker-assisted pyramiding of the three major genes for blast resistance in rice. Theor Appl Genet 100:1121–1128

Hoshida H, Tanaka Y, Hibino T, Hayashi Y, Tanaka A, Takabe T, Takabe T (2000) Enhanced tolerance to salt stress in transgenic rice that overexpresses chloroplast glutamine synthetase. Plant Mol Biol 43:103–111

Huang W, Cui X, Tian Y, Lin M, Peng X (1994) Cloning of T7 lysozyme gene and construction of the vector for transgenic plants resistant to bacterial infection. Weishengwu Xuebao 34:261–265

Ito Y et al. (2006) Functional analysis of rice DREB1/CBF-type transcription factors involved in cold-responsive gene expression in transgenic rice. Plant Cell Physiol 47:141–153

Neuhans JM, Ahl-Goy P, Hinz U, Flores S, Meins FJ (1991) High-level expression of a tobacco chitinase gene in *Nicotiana sylvestris*: susceptibility of transgenic plants to *Cercospora nicotianae* infection. Plant Mol Biol 16:141–151

Jackson RE, Bradley JR, Van Duyn JW (2003) Field performance of transgenic cottons expressing one or two *Bacillus thuringiensis* endotoxins against bollworm, *Helicoverpa zea* (Boddie). J Cotton Sci 7:57–64

Jackson R.E. et al. (2004) Performance of feral and Cry1Ac-selected *Helicoverpa zea* (*Lepidoptera*: Noctuidae) strains on transgenic cottons expressing one or two *Bacillus thuringiensis* ssp kurstaki proteins under greenhouse conditions. J Entomol Sci 39:46–55

Kang JY, Choi HI, Im MY, Kim SY (2002) Arabidopsis basic leucine zipper proteins that mediate stress-responsive abscisic acid signaling. Plant Cell 14:343–357

Kasuga M, Liu Q, Miura S, Yamaguchi-Shinozaki K, Shinozaki K (1999) Improving plant drought, salt and freezing tolerance by gene transfer of a single stress-inducible transcription factor. Nat Biotechnol 17:287–291

Katiyar-Agarwal S, Agarwal M and Anil Grover (2003) Heat tolerant basmati rice engineered by overexpression of Hsp101 gene. Plant Mol Biol. 51:677–686

Kim S-R, Lee J, Jun S-H, Park S, Kang H-G, Kwon S, An G (2003) Transgene structures in T-DNA-inserted rice plants. Plant Mol Biol 52:761–773

Kim S, Kang JY, Cho DI, Park JH, Kim SY (2004) ABF2, an ABRE binding bZIP factor, is an essential component of glucose signaling and its overexpression affects multiple stress tolerance. Plant J 40:75–87

Kalpana et al. (2006) Engineering sheath blight resistance in elite indica rice cultivars using genes encoding defense proteins. Plant Sci 170:203–215

Kramer KJ and Muthukrishnan S (1997) Insect Chitinases:Molecular biology and potential use as biopesticides. Insect Biochem Molec. Biol. 27:887–900

Kumar KK, Poovannan K, Nandakumar R, Thamilarasi K, Geetha C, Jayashree N, Kokiladevi E, Raja JAJ, Samiyappan R, Sudhakar D and Balasubramanian P (2003) A high throughput functional expression assay system for a defence gene conferring transgenic resistance on rice against the sheath blight pathogen, *Rhizactonia solani*. Plant Science 165:969–976

Linn W, Datta K, Potrykus I, Muthukrishnan S, Datta SK (1995) Genetic enginering of rice for resistance to sheath blight. Bio/Tech 13:686–691

Liu D, Burton S, Glancy T, Li ZS, Hampton R, Meade T, Merlo DJ (2003) Insect resistance conferred by 283-kDa Photorhabdus luminescens protein TcdA in Arabidopsis thaliana. Nat Biotechnol. 21(10):1222–8

Maqbool SB, Christou P, Riazuddin S, Loc NT, Gatehouse AMR, Gatehouse JA (2001) Expression of multiple insecticidal genes confers broad resistance against a range of different rice pests. Mol Breed 7:85–93

Mcgaughey WH, Gould F and Gelernter W (1998). Bt resistance management. Nature Biotechnology, 16 (2):144–146

Maqbool SB, Husnain T, Riazuddin S, Masson L, Christou P (1998) Effective control of yellow stem borer and rice leaf folder in transgneic rice indica varieties Basmati 370 and M7 using the novel delta endotoxin cry2A Bt gene. Mol Breed 4:501–507

Maroof MAS, Jeong SC, Gunduz I, Tucker DM, Buss GR, Tolin SA (2008) Pyramiding of soybean mosaic virus resistance genes by marker-assisted selection. Crop Sci 48:517–526

Maruthasalam S, Kalpana K, Kumar KK, Loganathan M, Poovannan K, Raja JA, Kokiladevi E, Samiyappan R, Sudhakar D, Balasubramanian P (2007) Pyramiding transgenic resistance in elite indica rice cultivars against the sheath blight and bacterial blight. Plant Cell Rep. 26:791–804

Martinez M, Abraham Z, Gambardella M, Echaide M, Carbonero P, Diaz I (2005) The strawberry gene Cyf1 encodes a phytocystatin with antifungal properties. J Exp Bot. 56(417):1821–9

Matsumura T, Tabayashi N, Kamagata Y, Souma C, Saruyama H (2002) Wheat catalase expressed in transgenic rice can improve tolerance against low temperature stress. Physiol Plant 116:317–327

Mcganghey WH, Gould F, and Gelernter W (1998) Bt resistance management. Nature Biotechnology, 16(2):144–146

Mew T, Vera, Cruz CM and Medalla ES (1992) Changes in race frequency of *Xanthomonas oryzae* pv. *oryzae* in response to rice cultivars planted in Philippines. Plant disease 76:1029–1032

Muthukrishnan S, Liang GH, Trick HN and Gill BS (2001) Pathogenesis-related proteins and their genes in cereals. Plant Cell, Tissue and Organ Culture 64, 93–114

Narayanan NN, Baisakh N, Oliva NP, VeraCruz CM, Gnanamanickam SS, Datta K, Datta SK (2004) Molecular breeding: marker-assisted selection combined with biolistic transformation for blast and bacterial blight resistance in indica rice (cv. CO39). Mol. Breed. 14:61–71

Ohta M, Hayashi Y, Nakashima A, Hamada A, Tanaka A, Nakamura T, Hayakawa T (2002) Introduction of a Na^+/H^+ antiporter gene from Atriplex gmelini confers salt tolerance to rice. FEBS Lett 532(3):279–282

Osmond RIW, Hrmova M, Fontaine F, Imberty A and Fincher GB (2001) Binding interactions between barley thaumatin-like proteins and (1,3)-β-D-glucans. Eur. J. Biochem., 268(15): 4190–4199

Outchkourov NS, de Kogel WJ, Wiegers GL, Abrahamson M, Jongsma MA (2004) Engineered multidomain cysteine protease inhibitors yield resistance against western flower thrips (Frankliniella occidentalis) in greenhouse trials. Plant Biotechnol J. 2(5):449–58

Pardo JM, Reddy MP, Yang S, Maggio A, Huh G, Matsumoto T, Coca MA, Paino-D'Urzo M, Koiwa H, Yun D, Watad AA, Bressan RA, Hasegawa PM (1998) Stress signaling through a2./ calmodulin-dependent protein phosphatase calcineurin mediates salt adaptation in plants. Proc Natl Acad Sci 95:9681–9686

Powell KS, Gatehouse AMR, Hilder VA, Van Damme EJM, Peumans WJ, Boonjawat J, Horsham K and Gatehouse JA (1995) Different antimetabolic effects of related lectins towards nymphal stages of *Nilaparvata lugens*. Entomologia Experimentalis et Applicata 75:61–65

Punja ZK (2006) Recent developments toward achieving fungal disease resistance in transgenic plants. Can J Plant Pathol 28:S298–S308

Quimio CA, Torrizo LB, Setter TL, Ellis M, Grover A, Abrigo EM, Oliva NP, Ella ES, Carpena AL, Ito O et al. (2000) Enhancement of submergence tolerance in transgenic rice overproducing pyruvate decarboxylase. J Plant Physiol 156:516–521

Raman H, Read BJ, Brown AHD, Abbott DC (1999) Molecular markers and pyramiding of multiple genes for resistance to scald in barley. In: Australian barley technical symposium

Ranjekar PK, Hinge A, Hegde MV (2003) Decreased antioxidant enzymes and membrane essential polyunsaturated fatty acids in schizophrenia and bipolar mood disorder patients. Psychiatry Res 121:109–122

Ronald P (1997) Making Rice Disease Resistant. Scientific American, 277:100–105

Roush RT (1997) Bt-transgenic crops: just another pretty insecticide or a chance for a new start in resistance management? Pesticide Sci., 51(3):328–334

Ruud M, Bosch D, and Stiekema W (1999) Bacillus thuringiensis toxin mediated insect resistance in plants. Trends in Plant Science 4:9–13

Saijo Y, Hata S, Kyozuka J, Shimamoto K, Izui K (2000) Over-expression of a single Ca^{2+}-dependent protein kinase confers both cold and salt/drought tolerance on rice plants. Plant J 23:319–327

Sankula S et al. (2005) Biotechnology-derived crops planted in 2004 – bimpacts on US agriculture. National Center for Food and Agricultural

Sawahel W (2003) Improved performance of transgenic glycinebetaine-accumulating rice plants under drought stress. Biol Plant 47:39–44

Schuler TH, Poppy GM, Kerry BR, Denholm I (1988) Insect resistant transgenic plants. Trends Biotechnol., 16:168–175

Servin B, Martin OC, Mézard M and Hospital F (2004) Toward a Theory of Marker-Assisted Gene Pyramiding. Genetics, 168, 513–523

Sharma PN, Akihide T, Shigeo T, Naoki M, Chiharu N (2004) Marker-assisted pyramiding of brown planthopper (Nilaparvata lugens Stål) resistance genes *Bphl* and *Bph2* on rice chromosome 12. Hereditas 140:61–69

Shelton AM, Zhao J, Roush RT (2002) Economic, ecological, food safety and social consequences of the deployment of Bt transgenic plants. Ann Rev Entomol 47:845–881

Shinozaki K, Yamaguchi-Shinozaki K (2000) Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. Curr Opin Plant Biol 3(3):217–223

Shinozaki K, Yamaguchi-Shinozaki K (1999) Molecular responses to drought stress. In: Shinozaki K, Yamaguchi-Shinozaki K (eds) Molecular responses to cold, drought, heat and salt stress in higher plants. R.G. Landes Co., Austin, TX, pp 11–28

Shou HX, Bordallo P, Wang K (2004) Expression of the Nicotiana protein kinase (NPK1) enhanced drought tolerance in transgenic maize. J Exp Bot 55:1013–1019

Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. Science 270:1774–1778

Stewart SD, Adamczyk JJ, Knighten KS and Davis FM (2001) Impact of *Bt* cottons expressing one or two insecticidal proteins of *Bacillus thuringiensis* Berliner on growth and survival of noctuid (Lepidoptera) larvae. J Econ Entomol., 94:752–760

Su J, Wu R (2004) Stress-inducible synthesis of proline in transgenic rice confers faster growth under stress conditions than that with constitutive synthesis. Plant sci 166:941–948

Takakura Y, Ito T, Saito H, Inoue, T, Komari T, Kuwata S (2000) Flower-predominant expression of a gene encoding a novel class I chitinase in rice (*Oryza sativa* L.). Plant Mol. Biol. 42:883–897

Takesawa T, Ito M, Kanzaki H, Kameya N, Nakamura I (2002) Over-expression of ζ glutathione *S*-transferase in transgenic rice enhances germination and growth at low temperature. Mol Breed 9:93–101

Tamura T, Hara K, Yamaguchi Y, Koizumi N, Sano H (2003) Osmotic stress tolerance of transgenic tobacco expressing a gene encoding a membrane-located receptor-like protein from tobacco plants. Plant Physiol131:454–462

Tang K, Tinjuangjun P, Xu Y, Sun X, Gatehouse JA, Ronald PC, Qi H, Lu X, Christou P and Kohli P (1999) Particle-bombardment-mediated co-transformation of elite Chinese rice cultivars with genes conferring resistance to bacterial blight and sap-sucking insect pests. Planta 208:552–563

Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. Bio/Technol 7:257–264

Takakura Y, Ito T, Saito H, Inoue T, Komari T, Kuwata S (2000). Flower-predominant expression of a gene encoding a novel class I chitinase in rice (*Oryza sativa* L.). Plant Mol. Biol. 42:883–897

Thomashow MF (1999) Plant cold acclimation. Freezing tolerance genes and regulatory mechanisms. Annu Rev Plant Physiol Plant Mol Biol 50:571–599

Tu J, Zhang G, Datta K, Xu C, He Y, Zhang Q, Khush GS, Datta SK (2000) Field performance of transgenic elite commercial hybrid rice expressing Bacillus thuringiensis d-endotoxin. Nat Biotechnol 18:1101–1104

Uno Y, Furihata T, Abe H, Yoshida R, Shinozaki K, Yamaguchi- Shinozaki K (2000) Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. PNAS 97:11632–11637

Uchimiya H, Fujii S, Huang J, Fushimi T, Nishioka M and Kim KM (2002) Transgenic rice plants conferring increased tolerance to rice blast and multiple environmental stresses. Molecular Breeding 9:25–31

Van Loon LC, Van Strien EA (1999) The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. Physiol Mol Plant Pathol 55:85–97

Velazhahan R, Chen-Cole K, Anuratha CS, and Muthukrishnan S (1998) Induction of thaumatin-like proteins (TLPs) in Rhizoctonia solani-infected rice and characterization of two new cDNA clones. Physiol Plant 102:21–28

Viswanathan C, Zhu JK (2004) Molecular perspectives on cross-talk and specificity in abiotic stress signaling in plants. J Exp Bot 55:225–236

Wang J, Chen Z, Du J, Sun Y and Liang A (2005) Novel insect resistance in Brassica napus developed by transformation of chitinase and scorpion toxin genes. Plant Cell Rep. 24(9):549–55

Werner K, Friedt W, Ordon F (2002) Marker-assisted pyramiding of resistance genes against the barley yellow mosaic virus complex (Bammv, Baymv, Baymv-2). In: Plant, Animal & Microbe Genomes X Conference, January 12–16, 2002, USA

Winicov I (1998) New molecular approaches to improving salt tolerance in crop plants. Ann Bot 82:703–710

Williamson VM, Kaya HK. (2003) Sequence of a symbiont. Nat Biotechnol. 21(11):1294–5

Xiong L, Yang Y (2003) Disease resistance and abiotic stress tolerance in rice are inversely modulated by an abscisic acid-inducible mitogen-activated protein kinase. Plant Cell 5:745–759

Yamanouchi U, Yano M, Lin H, Ashikari M, Yamada K (2002) A rice spotted leaf gene, Spl7, encodes a heat stress transcription factor protein. PNAS 99 (11):7530–7535

Ye GY, Tu J, Hu C, Datta K, Datta SK (2001) Transgenic IR72 with fused Bt gene cry1A(b)/cry1A(c) from *Bacillus thuringiensis* is resistant against four lepidopteran species under field conditions. Plant Biotech 18:125–133

Yeo AR, Flowers TJ (1989) Selection for physiological characters – examples from breeding for salt tolerance. In: Flowers TJ, Jones MBJHG (eds) Plants under stress biochemistry, physiology and ecology and their application to plant improvement. Cambridge University Press, Cambridge, pp 217–234

Yun DJ, Ibeas JI, Lee H, Coca MA, Narasimham ML, Uesono Y, Hasegawa PM, Pardo JM, Bressan RA (1998) Osmotin, a plant antifungal protein; subverts signal transduction to enhance fungal cell susceptibility. Mol Cell 1:807–817

Zhang S, Song WY, Chen L, Ruan D, Taylor N, Ronald PC, Beachy RN, Fauquet CM (1998) Transgenic elite indica rice varieties, resistant to *Xanthomonas oryzae* pv.oryzae. Mol Breed 4:551–558

Zhao JZ, Fan YL, Fan XL. (1999) Evaluation of transgenic tobacco expressing two insecticidal genes to delay resistance development of Helicoverpa armigera. Chinese Science Bulletin, 44(20):1871–1874

Zhu JK (2002) Salt and drought stress signal transduction in plants. Ann Rev Plant Biol 53:247–273

Zhu JK (2001) Cell signaling under salt, water and cold stresses. Curr Opin Plant Biol 4:401–406

Chapter 8
Application of Molecular Markers for Breeding Disease Resistant Varieties in Crop Plants

Ana M. Torres

Abstract Introgression of desired genes using traditional protocols is lengthy and complicated. Since the advent of the first DNA markers, marker assisted selection (MAS) has been viewed as a promising approach to streamline resistance breeding. Molecular markers linked to resistance genes can obviate the need for testing to identify resistant individuals from early generations, leading to an effective improvement of the breeding procedure. Today, the most successful applications of MAS in plant breeding have been those for major disease resistance genes assisting backcrossing into elite cultivars and selecting alleles with major effects on high-value traits with relatively simple inheritance. By contrast, the approach has not been used as widely for the improvement of polygenic traits, due to the insufficient precision of QTL mapping techniques and the unreliable extrapolation of QTL information across multiple populations. Improvements in marker technologies and major investments in economically important crops has revealed the full benefits of MAS in private companies. However, further genomic research and reductions in the costs associated with molecular markers are required to provide new opportunities to employ MAS in minor crops and public breeding programmes. This chapter provides an overview of the current status of MAS for breeding disease resistance in important food crops, and highlights strategies that will enhance the impact of MAS in the near future.

8.1 Introduction

Traditional protocols of plant breeding are based on the phenotypic selection of plants with traits of interest, with the final goal of assembling desirable combinations of genes in new varieties. These practices have been very effective in improving crop productivity during the past decades. However, conventional methods often

A.M. Torres
IFAPA, Centro Alameda del Obispo, Área de Mejora y Biotecnología,
Apdo. 3092, 14080, Córdoba, Spain
e-mail: anam.torres.romero@juntadeandalucia.es

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_8, © Springer Science+Business Media B.V. 2010

encounter difficulties related to genotype × environment interactions that can reduce the effectiveness of phenotypic selection and complicate the identification of superior genotypes. Moreover, screening protocols are expensive, time-consuming or often unreliable for particular traits such as disease resistance. Thus, the introgression of a resistance gene into a breeding line by traditional breeding is lengthy and complicated by the need of performing artificial inoculation tests to assess the resistant phenotype. This, in turn, requires maintenance of pathogens or pests on the host (or alternate hosts) if they are obligate parasites (Francia et al. 2005). As a consequence, the average length of a breeding program from hybridisation and selection of favourable genetic combinations to testing in the field and introduction into the market can vary from 10 to 15 years.

Plant breeders also must constantly respond to new challenges. Agricultural practices are changing, and crops are regularly exposed to altered growing conditions which create the need for developing genotypes with specific agronomic characteristics. Fungal and insect pests evolve continuously, overcoming host resistance and forcing plant breeders to the endless task of developing new crop varieties (Collard and Mackill 2008). Furthermore, genes for disease and pest resistance are often present only within wild species. Introgression into adapted cultivars is complicated by the potential loss of these genes during backcrosses and inbreeding processes to eliminate undesirable traits introduced from the wild donor (Foolad 2007). These and other problems associated with the use of traditional breeding methods demand the employment of techniques that have higher potential for resolution.

Molecular markers are considered valuable tools for crop improvement, due to their usefulness in characterizing and manipulating genetic loci responsible for monogenic and polygenic traits. Markers are helping researchers to understand complex traits, dissect them into single Mendelian components and establish their chromosomal locations using linkage maps and/or cytogenetic stocks. Availability of well saturated genetic linkage maps is a prerequisite for tagging traits with molecular markers, thus enabling their use in MAS (Marker Assisted Selection) related activities. The application of MAS to the introgression of genes from one genotype (the donor genotype) to another (the recurrent genotype) through a backcross breeding scheme clearly illustrates the great advantages of the use of molecular markers for indirect selection of traits that would otherwise be difficult to select for by standard procedures.

8.2 Molecular Marker Technologies

During the past two decades DNA marker technology has spurred tremendous advances in map development for most important food crops. It is beyond the scope of this chapter to discuss the technical details on how DNA markers are generated. However some advantages and disadvantages of the most commonly used marker types are exposed. For more extensive reviews on this topic see Joshi et al. (1999), Gupta et al. (2001), Khlestkina and Salina (2006), and Edwards and McCouch (2007) among many others.

The earlier types of molecular markers include anonymous or neutral markers based on hybridization, such as restriction fragment length polymorphisms (RFLPs), and which were later followed by based on the polymerase chain reaction (PCR), a faster and less expensive technology. PCR-based DNA markers include random amplified polymorphic DNAs (RAPDs), amplified fragment length polymorphisms (AFLPs), cleaved amplified polymorphic sequences (CAPS), and sequence characterized amplified regions (SCARs), which have been extensively used in different applications.

Even today, RFLPs offer the best marker type for many purposes. The main drawbacks of using RFLPs are the high cost and the low throughput of genotyping. RAPDs and AFLPs have also been widely used in genetic diversity studies and gene mapping. Both technologies are particularly useful when there is a need to assay loci across the entire genome. Nevertheless, their dominant nature, the lack of reproducibility of RAPDs compared with AFLPs and the lack of specificity in both cases, are limiting factors for their application in accurate MAS breeding approaches. However, random techniques such as RAPDs and AFLPs are highly useful for finding new markers linked to desirable alleles. Once such markers are identified, the corresponding bands can be sequenced and used to develop more specific and reliable markers such as CAPs or SCARs that simplify the screening of large progenies (Edwards and McCouch 2007).

A significant development in PCR marker technology is evident when the DNA sequence is available and it is possible to design primers to amplify across a highly variable locus. These highly variable features include tandem repeats such as microsatellites or SSRs, and dispersed complex repeats such as transposable elements. Microsatellites are relatively simple and cheap to use and have been employed for a multitude of genetic projects due to the highly reproducible and reliable identification of alleles.

In recent years, the availability of whole genome sequences of a few selected crops and the sequence information generated by expressed sequence tags (ESTs) has also led to the development of a new generation of markers, gene-targeted markers (also called candidate gene markers) and functional markers (Andersen and Lübberstedt 2003). These are derived from polymorphic sites within genes known to be causally involved in phenotypic trait variation. ESTs are currently the most widely sequenced elements from plant genomes and constitute a novel source of markers physically associated with coding regions, as well as a possible source of SSRs in many crops. ESTs thus provide a robust sequence resource that can be exploited for gene discovery, genome annotation, and comparative genomics.

The major drawback of anonymous markers with respect to functional markers is that their predictive value depends on a tight linkage between marker and target locus. Functional markers are, however, superior to anonymous markers due to their complete linkage with genes of known function. Once genetic effects have been assigned to a sequence motif, derived markers can be used efficiently to fix the corresponding alleles in a number of genetic backgrounds.

Gene-targeted markers and functional markers are often based on the discovery of single nucleotide polymorphisms (SNPs) between alleles. SNPs provide the

most abundant source of sequence variants encountered in most genomes (Cho et al. 1999; Picoult-Newberg et al. 1999), and are often the only option for finding markers that are very close to or within a gene of interest. Their development costs are similar to those of SSRs, but there is a myriad of SNP assay technologies which constitute some of the most highly automated, efficient and relatively inexpensive genotyping methods (Henikoff and Comai 2003; Kwok and Chen 2003).

8.3 Molecular Markers in Breeding Applications

The wide range of markers currently available has dramatically increased our knowledge of the genetic diversity within many plant species, and has greatly facilitated mapping of genomic regions that contribute to trait variation. Using the marker maps, putative genes affecting traits of interest have been detected by testing for statistical associations between marker variants and traits (Paterson et al. 1991). Following their identification, useful genes or quantitative trait loci (QTLs) can be introgressed into desirable genetic backgrounds via MAS, using markers physically located close to or even within genes of interest.

The potential of MAS as a tool for crop improvement has been extensively explored (Tanksley et al. 1989; Ribaut et al. 2002; Servin et al. 2004). MAS offers promise for:

1. Early screening of genotypes in the seedling stage, important for traits that are expressed late in the life cycle of the organism
2. Screening for rare recombinants between closely linked genes
3. Effective screening for traits that are extremely difficult, expensive or time consuming to score phenotypically
    a. Indirect selection of desirable plants avoiding environmental, pleiotrophic or epistatic effects
    b. Discriminating between homo- and heterozygous individuals in a single generation without the need for progeny testing
    c. Monitoring single or multiple trait/QTL introgression in backcrossing programs (known as gene pyramiding)

Holland (2004) described three levels of marker-assisted backcrossing. In the first level, markers are used in combination with or instead of screening for the target gene or QTL. This may be especially useful for selecting recessive alleles or for traits whose phenotypic screening protocols are more laborious or expensive than marker assays. The second level refers to the use of markers for selection of recombinants with chromosome segments containing the target gene and as little as possible surrounding DNA from the donor parent. This is critical to avoid undesirable genes that may negatively affect crop performance which is referred to as “linkage drag” (Hospital 2005). Finally, the third level of marker-assisted backcrossing involves using markers unlinked to the target locus for selection of recombinants with the

highest proportion of the recurrent parent (Hospital and Charcosset 1997; Frisch et al. 1999; Collard and Mackill 2008).

The MAS approach is not only a tool of speeding up the process of gene transfer, but also allows pyramiding of desirable genes and QTLs from different genetic backgrounds. Gene pyramiding refers to the introgression of several characters at one time, a method that may enhance the durability or degree of pest and disease resistances. When a cultivar is protected by one major gene it is difficult to introgress additional genes conferring resistance to the same disease because of the difficulty of their discrimination in phenotypic screenings (since the plant already shows resistance). However, by tagging several genes with closely linked molecular markers, MAS strategies facilitate the development of lines with stacked resistance genes, giving the cultivar more durable protection than that afforded by a single resistance gene (William 2007a). Also, genes controlling resistance to different races or biotypes of a pest or pathogen, or genes contributing to agronomic or seed quality traits can be pyramided together to maximize the benefit of MAS through simultaneous introgression (reviewed in Dwivedi et al. 2007).

The following section will focus on selected research relating to the identification, tagging, mapping, and MAS of resistance genes and QTLs in important food crops. Examples of marker-trait associations are too numerous to review in a single chapter, but we will present an overview of novel and successful applications of MAS in breeding for disease resistance and highlight the strategies to enhance the impact of MAS in the near future.

8.4 Efficient Applications of MAS in Breeding for Disease Resistance

Development of cultivars with improved resistance to biotic stresses is a primary goal of plant breeding programs throughout the world. Disease resistant cultivars can reduce both the risk of yield loss and the dependence on pesticides in high input systems, enabling a more stable production across diverse and adverse environments and soil conditions (Miklas et al. 2006).

Since the advent of the first DNA markers, MAS has been viewed as a promising approach to streamline resistance breeding. Today, the most successful applications of MAS in plant breeding have been those for major disease resistance genes. Molecular markers linked to resistance genes can obviate the need for resistance testing to identify resistant individuals in early generations of breeding populations, reducing the number of progeny maintained. Molecular markers are now routinely used in plant cultivar development to assist backcrossing of major genes into elite cultivars and to select alleles with major effects on high-value traits with relatively simple inheritance.

Although the use of MAS is most straightforward for manipulating single-gene traits, its potential for breeding complex traits also has been recognized (Bouchez

et al. 2002; Lecomte et al. 2004, among others). Excellent reviews on MAS applications in different crops have been provided by Gupta and Varshney (2004), Francia et al. (2005), Foolad (2007), Landjeva et al. (2007), Dwivedi et al. (2007) and Collard and Mackill (2008), among others. However, it should be noted that despite hundreds of published QTL studies, MAS for polygenic trait improvement is just now starting to produce convincing results, and relatively few practical applications have yet been published. This paucity of reports highlights the long term nature of this research, or according to Hoisington and Melchinger (2004), might reflect the fact that marker technology has been applied to plant breeding efforts mostly by scientists working in industry. Additional factors that are delaying the scope and utility of MAS in plant breeding will be exposed in Section 5, however the insufficient precision of QTL mapping techniques and unreliable extrapolation of QTL information across multiple populations are some of the main reasons.

Likewise, the availability of well-saturated linkage maps and the extent of genetic studies vary among different crops and also influence the feasibility of any MAS-related activity. Thus, these resources are available for MAS cereals such as rice, maize and wheat, and in legumes such as soybean. For other, lower value, species considered as "orphan crops", genetic improvement with MAS is not yet feasible. In this section, we briefly summarize selected cases where MAS has been used to incorporate disease resistance traits into improved genetic backgrounds of major food crops. The space limitation of the chapter does not allow for discussion of procedures or methodologies used, however, some general comments can be made as to the mapping and tagging of specific resistance genes and QTLs.

8.4.1 Cereals

As the most important components of daily diet for human population, cereals have received by far the strongest support from basic research. Cereal genomics has undergone exponential growth in the last two decades leading to the availability of a large number of molecular markers, well saturated genetic and physical maps and batteries of expressed sequenced tags (ESTs) and sequencing regions carrying specific genes. All these approaches are facilitating studies on comparative mapping and helping cereal breeding programs through MAS.

8.4.1.1 Rice

Rice was the first cereal to be fully sequenced because of its agricultural importance and its small genome size, low number of chromosomes, well characterized genetic and genomic resources, and availability of a large number of DNA markers and high-density genetic linkage maps (revised in Gowda et al. 2003). Consequently, the status of rice as a genomic model is set to promote the application of MAS in breeding.

Much of the progress to date has centred on marker-assisted backcrossing or pyramiding of genes against bacterial blight caused by *Xanthomonas oryzae* pv. *oryzae* and rice blast caused by the fungus *Magnaporthe grisea*. The International Rice Research Institute (IRRI, Philippines and Punjab Agricultural University, Ludhiana) have successfully employed MAS to pyramid genes for bacterial blight resistance. All possible combinations of the four resistance genes, *Xa4*, x*a5*, x*a13* and *Xa21* have been pyramided using RFLP and STS (sequence tagged site) markers (Sanchez et al. 2000; Chen et al. 2000, 2001; Singh et al. 2001, among others). The pyramided lines show a wider spectrum or higher level of resistance to the bacterial pathogen. Recently, the Government of Indonesia released two new rice varieties obtained through MAS, 'Angke' and 'Conde', which are resistant to bacterial leaf blight infection (after Babu et al. 2004). MAS has been used successfully used in gene pyramiding and nine varieties have been released, two each in Philippines and India, and five commercial hybrids in China. Rice blast caused by the fungal pathogen *Magnaporthe grisea* is another important disease. Three genes, namely *Pi1*, *Piz5* and *Pita* have been pyramided using RFLP and PCR-based markers for durable blast resistance in high-yielding rice cultivars (Hittalmani et al. 2000; Liu and Anderson 2003). By MAS has been used to transfer Bph 18 gene into commercial japonica cultivar in Korea (Jena et al. 2006).

Additional targets in rice for genetic enhancement through MAS include resistance to gall midge and to brown plant hopper, two of the most serious insect pests of rice in Asia. Of the ten known gall midge resistance genes, eight have been tagged and mapped, and tightly linked markers are being used in gene pyramiding into elite rice cultivars (Mohan et al. 1994, 1997; Nair et al. 1995; Katiyar and Bennett 2001; Sardesai et al. 2002; Jain et al. 2004). Brown planthopper is a destructive phloem feeding insect pest against which both qualitative (Jena et al. 2003, 2006) and quantitative resistance (Xu et al. 2002) has been reported. An example of gene pyramiding was reported by Sharma et al. (2004) who constructed a gene-pyramided *japonica* line containing two brown plant hopper resistance genes, *Bph1* and *Bph2*, derived from two *indica* resistance lines.

8.4.1.2 Barley

Selected examples of successful MAS application in barley include the early identification of homozygous or heterozygous plants resistant against barley mild mosaic virus (BaMMV) and barley yellow mosaic virus BaYMV (Tuvesson et al. 1998). There are numerous resistance sources against these viruses, but resistance in cultivars of European winter barley depends entirely on a single recessive gene (*ym4*), which confers complete immunity. One RFLP marker tightly linked to *ym4* (Graner and Bauer 1993) was converted into a PCR-based STS marker (Bauer and Graner 1995). Based on this marker, plants possessing homozygous or heterozygous resistance against BaMMV/BaYMB are being used actively in commercial breeding programmes. A variety of different markers have been developed in DH populations for selection of *rym4* and the allelic *rym5* resistance genes (Graner et al. 1999). One of them, the SSR Bmac0029, is used at present by many European

winter barley breeders (Rae et al. 2007). Results from genomics studies have further lead to the identification of a candidate gene for resistance to the Yellow Mosaic Virus complex (Wicker et al. 2005). Finally, the fine-mapping (Pellio et al. 2005) and recent cloning of the *rym4/5* locus (Stein et al. 2005) opens up the prospect of a diagnostic marker for *rym4/5*-based virus-resistance (reviewed in Rae et al. 2007). Further molecular markers available in barley breeding for virus resistance have been reported by Ordon et al. (2004).

An example of pyramiding in barley involves stacking of the resistance genes *rym4, rym5, rym9* and *rym11* for the barley yellow mosaic virus complex using molecular markers and doubled haploids (Werner et al. 2005). Additional markers have been reported for barley yellow dwarf virus (*Yd2*) (Jefferies et al. 2003) and leaf rust (*Rphq6*) (van Berloo et al. 2001), broadening the number of resistance genes that may be combined in a barley MAS programme (Rae et al. 2007; Mammadov et al. 2007).

Barley stripe rust caused by *Puccinia striiformis* f. sp. *hordei* is another disease that causes severe quality and yield losses. Toojinda et al. (1998) described the use of RFLPs to introgress the stripe rust resistance QTL in elite breeding material through marker-assisted backcrossing. The procedure has allowed the rapid development of resistant germplasm of barley on a commercial scale. The US variety “Tango” is claimed to be the first commercially released cultivar obtained by molecular MAS.

8.4.1.3 Wheat

Linkage map construction in wheat is more challenging than in species such as rice or maize due to its hexaploid nature and a large genome size composed of 21 linkage groups, as opposed to 10 in maize and 12 in rice. The use of MAS in wheat has increased over the last few years, encouraged by a surge in the number of amenable target traits and by a combination of technological improvements. In 2001, a US wheat MAS consortium involving public sector breeders was formed with the aim to transfer developments in wheat genomics to wheat production (Dubcovsky 2004). The program targeted traits such as disease and pest resistance or end-use quality, and was seeking the transfer of these traits through MAS. Eagles et al. (2001) reported a parallel use of a number of marker assays in wheat cultivar development programs in Australia. In public wheat breeding programs in Canada, MAS is an integrated component for traits such as midge and rust resistance (William 2007a). Finally, European wheat breeding companies are also known to use markers in their breeding efforts (Angus 2007), but unfortunately no data is publically available.

The most appropriate targets for MAS in current wheat breeding programs are Fusarium head blight, which is difficult to evaluate, rust and virus resistance. Successful examples of single gene introgression and gene pyramiding include the genetic enhancement of resistance to leaf rust through the genes *Lr19, Lr51* and *Yr15* (Slikova et al. 2003; Helguera et al. 2005; Chagué et al. 1999) and the development of improved pyramided lines and cultivars containing combinations for the powdery mildew resistance genes *Pm2, Pm4a, Pm6, Pm8* and *Pm21*

(Liu et al. 2000; Wang et al. 2001), as well as lines with resistance to Fusarium head blight, orange blossom midge (*Sm1*), and leaf rust (*Lr21*) (Somers et al. 2005).

Another good candidate for widespread use of MAS in wheat breeding programs is resistance to the cereal cyst nematode *Heterodera avenae*. Recently, Barloy et al. (2006) reported the pyramiding, through MAS, of two resistance genes, *CreX* and *CreY*, from the wild grass species *Aegilops variabilis* into a wheat background. Moreover, a novel QTL controlling resistance to the cereal cyst nematode in wheat has recently been associated with microsatellite marker loci (Williams et al. 2006). The locus acts additively with previously identified resistance loci *Cre8* and *Cre5*.

MAS is being used at CIMMYT to facilitate selection of a set of traits with low heritability, expensive and time-consuming screening protocols, but high economic value. A set of markers linked to genes controlling resistance to foliar diseases, root health, and quality are routinely used for their introgression into high-yielding backgrounds (William et al. 2007b). Examples of the diagnostic or perfect markers that are currently in use are cereal cyst nematode resistance genes *Cre*1 and *Cre*3 (Lagudah et al. 1997), barley yellow dwarf virus resistance introgressed from *Thinopyrum intermedium* on chromosome 7DL (Ayala et al. 2001), and a marker for *Aegilops ventricosa*-derived resistance to stripe rust (*Yr17*), leaf rust (*Lr37*) and stem rust (*Sr38*). CIMMYT also uses a set of linked markers for transferring a locus with major effects on boron tolerance (*Bo-1*), crown rot resistance, scab resistance and stem rust resistance (reviewed by William et al. 2007b).

8.4.1.4 Maize

Together with rice, maize was one of the first crop species for which molecular linkage maps were developed and merged into a consensus map (Gardiner et al. 1993). Maize is a naturally out-crossing species that shows highly significant levels of heterosis. In contrast to the situation in the other major cereals, where breeding is commonly carried out by public organizations, maize breeding in developing countries is controlled by private sector entities. Maize production in industrialized countries is also focussed towards the development of F1 hybrids and dominated by a small number of multinational private sector companies that are able to sustain profitability through their control over the genotype of their varieties. This has far-reaching implications on the feasibility of MAS in maize, and largely explains the lead that maize enjoys over wheat and other minor cereals in the deployment of MAS technology (Koebner and Summers 2007).

The most common use of MAS by the private sector is related to backcrossing of maize transgenes for resistance to herbicides or insects into elite inbred lines, the direct parents of the commercial hybrids (Ragot et al. 1995; Crosbie et al. 2006). To satisfy regulatory requirements and market demands, companies normally develop parallel breeding programmes with non-transgenic maize line. MAS schemes and infrastructure have recently facilitated the introgression of native

genes and QTLs for relatively complex traits such as resistance to grey leaf spot, northern corn leaf blight, Fusarium stalk and ear rots, abiotic stress tolerance and grain yield (Koebner and Summers 2007). The prospect of MAS to improve insect resistance in tropical maize has been less efficient than conventional phenotypic selection (Bohn et al. 2001). However, an increase in the relative efficiency is achieved when MAS and conventional phenotypic screening are combined (Collard and Mackill 2008).

Selected examples of MAS include resistance against northern corn blight resistance and southwestern corn borer (Simcox and Bennetzen 1993; Willcox et al. 2002, respectively). William et al. (2007a) highlighted CIMMYT experience with MAS applications in maize. For example, a major QTL controlling maize streak virus resistance explaining 50–70% of total phenotypic variation (Pernet et al. 1999a, b). Later, several microsatellite markers associated with this QTL were successfully used for the selection of resistant lines (William et al. 2007a).

To facilitate marker application in public efforts to improve wheat and maize cultivars, CIMMYT has recently established a marker implementation laboratory that provides the facilities and technical expertise for its breeders. The laboratory carries out two main MAS-related activities (1) identification in the literature of markers developed by third parties and verification that these can be used to detect traits or genes of interest in CIMMYT germplasm; (2) growth and/or sampling of plant tissue, DNA extraction, marker detection, data analysis and dissemination of results to breeders (William et al. 2007b).

8.4.2 Legumes

In contrast to cereals, there are very few reports in legumes of successful single gene transfer by MAS, and these are limited to two crops, soybean and common bean. This difference is due to the early stage of genomic research in these crops and to the limited number of trait mapping studies that have been completed (Dwivedi et al. 2007).

Loci for resistance to cyst nematode in soybean and common bacterial blight in common bean have been transferred into improved breeding lines (Concibido et al. 1996; Yu et al. 2000). In Resistance to soybean cyst nematode, phenotypic assays for resistance screening take approximately 5 weeks and extensive greenhouse space and labour. Identification of closely linked microsatellite markers has enabled the efficient transfer of the resistance gene *rhg1* (Cregan et al. 2000). A second example of MAS in common bean was reported by Yu et al. (2000), who used SCARs associated with common bacterial blight resistance for efficient discrimination of resistant genotypes.

Reports on successful gene pyramiding in legumes include combining a QTL for resistance to corn earworm and *Pseudoplusia includens* (soybean looper) in a line expressing a synthetic *Bacillus thuringiensis cry1Ac* transgene in soybean (Walker et al. 2002, 2004). In common bean, resistance genes for rust, caused by *Uromyces*

appendiculatus, and anthracnose, caused by *Colletotrichum lindemuthianum*, have been pyramided in adapted cultivars through marker-assisted backcrossing (Faleiro et al. 2004). Resistances to common bacterial blight, bean common mosaic virus, and anthracnose have been also combined (Kelly et al. 2003).

Numerous examples of gene-marker associations exist in other minor legume crops such as pea, chickpea or faba bean. RAPDs and SCARs linked to genes controlling powdery mildew resistance in pea (Timmerman et al. 1994; Tiwari et al. 1998; Fondevilla et al. 2008a) are available and could be utilized in MAS strategies. Moreover, QTLs for Ascochyta blight resistance have identified and validated (Timmerman–Vaughan et al. 2004; Fondevilla et al. 2008b) and candidate genes for the quantitative response have been reported (Prioul-Gervais et al. 2007), which may enable breeders to rapidly select for resistance to this pathogen.

In chickpea, efforts to employ MAS have been initiated and resistance to fusarium wilt may be combined with other desired traits with relative ease. Ascochyta resistance, however, is more unclear despite the fact that QTLs for resistance have been localized at a few loci in the chickpea genome (reviewed in Millán et al. 2006). In faba bean, molecular breeding for resistance to broomrape (*Orobanche crenata*), *Ascochyta fabae* and rust (*Uromyces viciae-fabae*) has been carried out, and promising results for MAS applications have been obtained (Torres et al. 2006 and references therein).

8.4.3 Solanaceae

8.4.3.1 Tomato

Tagging and mapping of single-gene traits in tomato started much earlier than in other crop species. As a result, tomato is very rich in available molecular markers. Most of them, including RFLPs and PCR-based markers, however, fail to detect polymorphisms within the cultivated species or between the cultivated and closely related wild species such as *L. pimpinellifolium*, restricting their use in breeding programs. Most recently significant efforts have been devoted to the discovery of high-resolution genetic markers such as SNPs and insertion deletions (InDels) (Labate and Baldo 2005).

Mapping disease resistance genes and QTLs has been the focus of many mapping activities in tomato although most of these activities are not reported in public literature. Since the 1970s, with the pioneering work of Rick and Fobes (1974) identifying an association between root-knot nematode resistance and the isozyme locus *Aps-1* (acid phosphatase), MAS has become a reality for seed companies selecting for nematode resistance. A survey by Foolad (2007) of some major tomato seed companies in the United States and in Europe indicated that MAS is routinely employed for tomato improvement for many qualitatively inherited disease resistance traits. Examples include vertical resistance to diseases such as corky root, fusarium wilt, late blight, root-knot nematodes, powdery mildew, bacterial speck, tobacco/tomato mosaic virus, tomato spotted wilt virus, and verticillium wilt. Many companies

indicated as well that for several of the resistance traits, MAS was not only faster than phenotypic selection but in some cases also cheaper and more effective.

On the other hand, the author obtained very few examples of seed companies successfully using MAS for manipulating QTLs, although it was apparently attempted for bacterial wilt, bacterial canker, bacterial wilt, powdery mildew and yellow leaf curl virus. In public tomato breeding programs the use of MAS is less common, although it has been practiced to improve vertical resistance to some diseases such as late blight, bacterial canker (Coaker and Francis 2004), bacterial speck and bacterial spot (Yang and Francis 2005) and horizontal resistance to blackmold and late blight (Robert et al. 2001; Brouwer and Clair 2004). For an extensive review on these topics see Foolad (2007).

8.4.3.2 Potato

Currently, the potato map is among those most highly saturated with different molecular markers, thus providing extensive opportunities for optimal application of MAS schemes. There are more than 20 single dominant resistance genes (R genes) mapped on 10 chromosomes (reviewed in Gebhardt and Valkonen 2001), together with genes controlling quantitative resistance traits. The first QTLs mapped were those conferring resistance to insects (Bonierbale et al. 1994; Yencho et al. 1996), followed by QTLs for resistance to *Phytophthora infestons, Erwinia carotovora*, and *Globodera* spp. (Gebhardt and Valkonen 2001) and to potato leafroll virus (Marczewski et al. 2001).

A fine example of MAS was reported by Hämäläinen et al. (1997), who were able to screen many diploid and tetraploid Solanum genotypes for resistance to potato virus Y (PVY, genus Potyvirus) thanks to one RFLP marker linked to the gene Ry_{adg}. CAPs and SCARs have allowed the breeding of genotypes resistant to PVY (Kasai et al. 2000) and their use for the selection of quantitative traits, such as resistance to *Phytophthora infestans*, has also been suggested (Oberhagemann et al. 1999).

Reports on gene pyramiding in potato include the use of four PCR-based diagnostic assays to combine the Ry_{adg} gene for extreme resistance to PVY with *Gro1* for nematode resistance and with *Rx1* for extreme resistance to potato virus X (PVX, genus Potexvirus), or with *Sen1* for wart resistance (*Synchytrium endobioticum*). Marker-based selection of tetraploid potato clones showed multiple resistances to the four diseases (Gebhardt et al. 2006).

8.5 Future Challenges and Perspectives for MAS

Most traits of agronomic importance, including disease resistance, are difficult to breed due to their genetic complexity, the interactions between the underlying genes (epistasis) and their environment-dependent expression. The difficulty of phenotypic selection of these quantitative traits gave rise to an optimistic view of

the prospects of MAS for QTL selection. However, to date, few studies have demonstrated the advantages of marker-QTL associations over phenotypic selection in the development of enhanced breeding materials through introgression or gene pyramiding. Selected examples of success of MAS for quantitative traits have been provided by Ahmadi et al. (2001) in rice, Yousef and Juvik (2001) in maize, Walker et al. (2002, 2004) in soybean, Somers et al. (2005) in wheat and Richardson et al. (2006) in barley. For a more extensive review see Dwivedi et al. (2007).

Molecular markers should have a much greater impact on plant breeding if they could be used widely to aid selection of quantitative traits in a time and cost-effective manner. However, despite the advances in DNA marker development, the construction of high density genetic linkage maps and the extensive QTL mapping of economically important traits, MAS has been much more effective for relatively simple traits than for complex ones. Currently, the access to more reliable PCR-based markers such as SSRs and SNPs, and the recent development of gene-targeted and functional markers is facilitating mapping of QTLs with large effects and with a high level of resolution (Paterson et al. 1988). However, as suggested earlier by Young (1999), research on quantitative traits has several important challenges:

1. Improvement of scoring methods. The accuracy of phenotypic evaluation is of utmost importance for the accuracy of QTL mapping. For this reason, establishing reliable inoculation and scoring methods are continuous challenges for plant pathologists and breeders.
2. Use of larger population sizes to avoid the bias in estimates of QTL effects (Melchinger et al. 1998). The larger the population, the more accurate the mapping study and the more likely it allows detection of QTLs with smaller effects An increase in population size provides gains in statistical power, estimates of gene effects and confidence intervals of the locations of QTLs (Collard et al. 2005).
3. Validation of QTLs in multiple replications and environments. Quantitative traits often have a low heritability, with many QTLs segregating for the trait, each with small individual effects. As a result, effects of individual regions are not easily identified and multiple genomic regions must be manipulated at the same time in order to have a significant impact. Although the effects of many QTLs seem to be consistent across environments, the magnitude of their effects varies depending on environmental conditions, due to QTL × environment interactions. For this reason, replicates of field tests are required to accurately characterize the effects of QTLs and to evaluate their stability across environments. These interactions remain a major constraint for the discovery of QTLs that confer a consistent advantage across a wide range of environments and should therefore be carefully considered in developing an effective MAS scheme (Francia et al. 2005; Collard and Mackill 2008).
4. The effect of a QTL can depend on the genetic background. This emphasizes the importance of testing the QTL effects and the reliability of marker associations in various genetic backgrounds and, whenever possible, in parallel populations, before MAS is undertaken.
5. Improvement of the predictive and/or diagnostic values of markers. In case of markers linked to the QTL, the proportion of total phenotypic variance explained

by each QTL is key to the value of the marker in enhancing the breeding gain for the target trait. Similarly, the process of marker validation is required to determine the reliability of a marker (or flanking markers) to predict phenotypes.

All these approaches should increase the reliability and accuracy of QTL mapping studies. However, there are other potential limitations to the widespread use of MAS in plant breeding. Collard and Mackill (2008) indicate the existence of "physical gaps" between research laboratories and breeding institutes or companies, which is obvious when these groups are independent or do not work closely together. The situation is aggravated by the low motivation within academic institutions to ensure development of user-friendly markers which are validated in different situations and efficiently applied in breeding programmes, once the innovative research is published. In case of private companies, transfer of markers and relevant information to breeders is practically zero, since publication of results is generally discouraged.

The same author pointed out the "knowledge gaps" between molecular biologists, plant breeders and other related areas, due to the rapid developments of fundamental concepts and methodologies used by each of the disciplines in the past two decades. Both factors are restricting the integration between conventional and molecular breeding, limiting the transfer of markers and relevant information for breeding applications and ultimately affecting the development of breeding lines.

Apart from these reasons, one crucial limitation for the application of MAS in minor crops is the high cost of DNA assays compared with conventional phenotypic selection (Koebner and Summers 2002, 2003; Bonnett et al. 2005). As stated by Dekkers and Hospital (2002) economics is a key factor in the application of molecular breeding approaches and their use will be determined by the economic benefit relative to conventional selection. MAS is justified when it replaces more expensive or tedious assays, or results in an increased precision in the identification of the desired genotypes.

For economically important crops such as maize and soybean, which are controlled by large private sector companies, major investments have been made to develop and use novel marker technologies and associated assay platforms. These efforts have resulted in increased efficiency, higher throughput, and lower assay costs, enabling markers to be used extensively in MAS approaches. Therefore, it is desirable to consider MAS approaches on a case-by-case basis, taking into account the importance of a trait in the overall breeding scheme, the available resources in personnel and consumables, and the nature of the breeding material (William et al. 2007a). Moreover integration of genomics with molecular breeding for disease resistance particularly gene based MAS should offer a comprehensive research strategy for more efficient selection in the near future. Currently, a variety of high-throughput genotyping technologies are becoming sufficiently cheap to allow their broad use in plant breeding. A new generation of molecular markers based on the detection of SNPs promises high-throughput assays at relatively low costs, along with the potential for high levels of multiplexing (Francia et al. 2005). Moreover, it should be considered that, although in marker assisted backcrossing the initial cost of using markers can be more expensive compared to traditional breeding, in the

medium to long term time savings could lead to an accelerated variety release which in turn produces greater gains.

At present, traditional breeding efforts are being greatly enhanced through the integration of comparative genomics. Gene similarities and synteny across genomes mean that much of the information generated on any plant species has relevance to other plant species. Most of the economically important species of grasses, legumes and solanaceae have detailed comparative maps allowing prediction of gene content and gene order across species (Tanksley et al. 1992; Devos and Gale 1997; Kalo et al. 2004; Choi et al. 2004). The information from synteny studies on genomic regions bearing traits of interest in model or economically important crops can be directly applied to related species for the improvement of the trait.

Large scale sequencing projects have produced an exponentially growing number of ESTs that have been used for the development of marker-dense transcriptional maps in different species (Choi et al. 2007; Kota et al. 2007; Schneider et al. 2007). This has two important implications for the improvement of complex traits: (1) the possibility of exploiting the information derived from syntenic regions to related species and (2) using the maps as a source of candidate genes for QTL mapping. As stated by Grattapaglia (2007), the generalized use of an increasing set of interspecific transferrable markers and the derived consensus mapping information will allow faster and more detailed studies on QTL synteny among species. It will also facilitate validation of QTLs and expression-QTLs across variable genetic backgrounds, and positioning of a growing number of candidate genes co-localized with QTLs, to be tested in association mapping experiments.

A key step of the control of stress responses in plants is the transcriptional activation or repression of genes. Identification of differentially expressed genes is particularly important to understand stress responses in plants (Dita et al. 2006). Sequence information, while valuable and necessary as a starting point, is insufficient to answer questions concerning gene function. Recent advances in functional genomics are helping to assess gene function through genome-wide experimental approaches. Innovative tools such as DNA microarrays together with ESTs are emerging technologies in crop improvement allowing the quantitative assessment of RNA levels in the sample. This in turn should assist the breeder in selecting the best breeding lines, based on RNA expression profiles, as much as on marker genotypes (Babu et al. 2004). The combination of greater knowledge and new tools will lead to changes in the strategies used for breeding. Integrating genomics with molecular breeding for disease resistance particularly gene based MAS should offer a comprehensive research strategy for more efficient selection in the near future.

References

Ahmadi N, Albar L, Pressoir G et al. (2001) Genetic basis and mapping of the resistance to rice yellow mottle virus. III. Analysis of QTL effciency in introgressed progenies confirmed the hypothesis of complementary epistasis between two resistance QTL. Theor Appl Genet 103:1084–1092

Andersen JR, Lübberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554–560

Angus WJ (2007) The role of marker assisted selection in a commercial wheat breeding programme. Plant and Animal Genome XV, San Diego, CA, USA, p 56
Ayala L, Henry M, Gonzalez-de-Leon D et al. (2001) A diagnostic molecular marker allowing study of Th. Intermedium derived resistance to BYDV in bread wheat segregating populations. Theor Appl Genet 102:942–949
Babu R, Nair SK, Prasanna BM, Gupta HS (2004) Integrating marker-assisted selection in crop breeding – prospects and challenges. Curr Sci 87(5):607–619
Barloy D, Lemoine J, Abelard P et al. (2006) Marker-assisted pyramiding of two cereal cyst nematode resistance genes from *Aegilops variabilis* in wheat. Mol Breed 20:31–40
Bauer E, Graner A (1995) Basic and applied aspects of the genetic analysis of the *ym4* virus resistance locus in barley. Agronomie 15:469–473
Bohn M, Groh S, Khairallah MM (2001) Re-evaluation of the prospects of marker-assisted selection for improving insect resistance against *Diatraea* spp. in tropical maize by cross validation and independent validation. Theor Appl Genet 103:1059–1067
Bonierbale MW, Plaisted RL, Pineda O et al. (1994) QTL analysis of trichome-mediated insect resistance in potato. Theor Appl Genet 87:973–987
Bonnett DG, Rebetzke GJ, Spielmeyer W (2005) Strategies for effcient implementation of molecular markers in wheat breeding. Mol Breed 15:75–85
Bouchez A, Hospital F, Causse MA et al. (2002) Marker-assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. Genetics 162(4):1945–1959
Brouwer DJ, St. Clair DA (2004) Fine mapping of three quantitative trait loci for late blight resistance in tomato using near isogenic lines (NILs) and sub-NILs. Theor Appl Genet 108(4):628–638
Chagué V, Fahima T, Dahan A et al. (1999) Isolation of microsatellite and RAPD markers flanking the *Yr15* gene of wheat using NILs and bulked segregant analysis. Genome 42(6):1050–1056
Chen S, Lin XH, Xu CG et al. (2000) Improvement of bacterial blight resistance of Minghui 63, an elite restorer line of hybrid rice by marker-assisted selection. Crop Sci 40:239–244
Chen S, Xu CG, Lin XH, Zhang Q (2001) Improving bacterial blight resistance of 6078, an elite restorer line of hybrid rice, by molecular marker-assisted selection. Plant Breed 120:133–137
Cho RJ, Mindrinos M, Richards DR et al. (1999) Genome-wide mapping with biallelic markers in *Arabidposis thaliana*. Nat Genet 23:203–207
Choi H-K, Mun J-H, Kim D-J et al. (2004) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci USA 101:15289–15294
Choi IY, Hyten DL, Matukumalli LK et al. (2007) A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. Genetics 176(1):685–696
Coaker GL, Francis DM (2004) Mapping, genetic effects and epistatic interaction of two bacterial canker resistance QTLs from *Lycopersicon hirsutum*. Theor Appl Genet 108(6):1047–1055
Collard BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Phil Trans R Soc B 363:557–572
Collard BCY, Jahufer MZZ, Brouwer JB et al. (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica* 142:169–196
Concibido VC, Denny RL, Lange DA et al. (1996) RFLP mapping and molecular marker-assisted selection of soybean cyst nematode resistance in PI209332. Crop Sci 36:1643–1650
Cregan PB, Mudge J, Fickus EW et al. (2000) Two simple sequence repeat markers to select for soybean cyst nematode resistance conditioned by the rhg1 locus. Theor Appl Genet 99:811–818
Crosbie TM, Eathington SR, Johnson GR et al. (2006) Plant breeding: past, present, and future. In: Lamkey KR and Lee M (eds) Plant breeding: The Arnel R. Hallauer International Symposium. Blackwell Publishing, Ames, IA, USA
Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. Nat Rev Genet 3:22–32
Devos KM, Gale MD (1997) Comparative genetics in the grasses. Plant Mol Biol 35:3–15
Dita MA, Rispail N, Prats E et al. (2006) Biotechnology approaches to overcome biotic and abiotic stress constraints in legumes. Euphytica (2006) 147:1–24

Dubcovsky J (2004) Marker assisted selection in public breeding programs: the wheat experience. Crop Sci 44:1895–1898
Dwivedi SL, Crouch JH, Mackill DJ et al. (2007) The molecularization of public sector crop breeding: progress, problems, and prospects. Adv Agron 95:163–318
Eagles HA, Bariana HS, Ogbonnaya FC et al. (2001) Implementation of markers in Australian wheat breeding. Aust J Agric Res 52:1349–1356
Edwards JD, McCouch SR (2007) Molecular markers for use in plant molecular breeding and germplasm evaluation. In: Guimarães E, Ruane J, Scherf B, Sonnino A, Dargie J (eds) Marker-assisted selection. Current status and future perspectives in crops, livestock, forestry and fish. Food and Agriculture Organization of the United Nations, Rome
Faleiro FG, Ragagnin VA, Moreira MA et al. (2004) Use of molecular markers to accelerate the breeding of common bean lines resistant to rust and anthracnose. Euphytica 138(3):213–218
Fondevilla S, Rubiales D, Moreno MT et al. (2008a) Identification and validation of RAPD and SCAR markers linked to the gene *Er3* conferring resistance to *Erysiphe pisi* DC in pea. Mol Breed 22:193–200.
Fondevilla S, Satovic Z, Rubiales D et al. (2008b) Mapping of quantitative trait loci for resistance to *Mycosphaerella pinodes* in *Pisum sativum* subsp. *syriacum*. Mol Breed 21:439–454
Foolad MR (2007) Genome Mapping and Molecular Breeding of Tomato. Int J Plant Genomics. Article ID 64358:52
Francia E, Tacconi G, Crosatti C et al. (2005) Marker assisted selection in crop plants. Plant Cell Tissue Org 82:317–342
Frisch M, Bohn M, Melchinger AE (1999) Minimum sample size and optimal positioning of flanking markers in marker-assisted backcrossing for transfer of a target gene. Crop Sci 39:967–975
Gardiner JM, Coe EH, Melia-Hancock S (1993) Development of a core RFLP map in maize using an immortalized F2 population. Genetics 134:917–930
Gebhardt C, Valkonen JPT (2001) Organization of genes controlling disease resistance in the potato genome. Annu Rev Phytopathol 39:79–102
Gebhardt C, Bellin D, Henselewski H et al. (2006) Marker-assisted combination of major genes for pathogen resistance in potato. Theor Appl Genet 112:1458–1464
Gowda M., Venu RC, Roopalakshmi K et al. (2003) Advances in rice breeding. Mol Breed 11:337–352
Graner A, Bauer E (1993) *Rflp* mapping of the *Ym4* virus resistance gene in Barley. Theor Appl Genet 86:689–693
Graner A, Streng S, Kellermann A et al. (1999) Molecular mapping and genetic fine-structure of the *rym5* locus encoding resistance to different strains of the Barley Yellow Mosaic Virus Complex. Theor Appl Genet 98:285–290
Grattapaglia D (2007) Marker-assisted selection in *Eucalyptus*. In: Guimarães E, Ruane J, Scherf B, Sonnino A, Dargie J (eds) Marker-assisted selection. Current status and future perspectives in crops, livestock, forestry and fish. Food and Agriculture Organization of the United Nations, Rome
Gupta PK, Varshney RK (eds) (2004) Cereal genomics. Kluwer Academic, Dordrecht, the Netherlands
Gupta PK, Roy JK, Prasad M (2001) Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Curr Sci 80:524–535
Hämäläinen JH, Watanabe KN, Valkonen JPT et al. (1997) Mapping and marker–assisted selection for a gene for extreme resistance to potato virus Y. Theor Appl Genet 94:192–197
Helguera M, Soria VM, Khan IA et al. (2005) PCR markers for *Triticum speltoides* leaf rust resistance gene *Lr51* and their use to develop isogenic hard red spring wheat lines. Crop Sci 45:728–734
Henikoff S, Comai L (2003) Single-nucleotide mutations for plant functional genomics. Ann Rev Plant Biol 54:357–401
Hittalmani S, Parco A, Mew TV et al. (2000) Fine mapping and DNAmarker-assisted pyramiding of the three major genes for blast resistance in rice. Theor Appl Genet 100:1121–1128

Hoisington DA, Melchinger AE (2004) From theory to practice – marker assisted selection in maize. In: Lorz H, Wenzel G (eds) Biotechnology in agriculture and forestry: molecular marker systems, Springer-Verlag, Berlin, Heidelberg, Germany, pp 335–352

Holland JB (2004) Implementation of molecular markers for quantitative traits in breeding programs-challenges and opportunities. In: New directions for a diverse planet. Proceedings of the 4th International Crop Science Congress, Brisbane, Australia

Hospital F (2005) Selection in backcross programmes. Phil Trans R Soc B 360:1503–1511

Hospital F, Charcosset A (1997) Marker-assisted introgression of quantitative trait loci. Genetics 147:1469–1485

Jain A, Ariyadasa R, Kumar A et al. (2004) Tagging and mapping of a rice gall midge resistance gene, *Gm8*, and development of SCARs for use in marker-aided-selection (MAS) and gene pyramiding Theor Appl Genet 109:1377–1384

Jefferies SP, King BJ, Barr AR et al. (2003) Marker-assisted backcross introgression of the *Yd2* gene conferring resistance to barley yellow dwarf virus in barley. Plant Breed 122:52–56

Jena KK, Pasalu IC, Rao YK et al. (2003) Molecular tagging of a gene for resistance to brown planthopper in rice (*Oryza sativa* L). Euphytica 129:81–88

Jena KK, Jeung JU, Lee JH et al. (2006) High-resolution mapping of a new brown planthopper (BPH) resistance gene, *Bph18*(t), and marker-assisted selection for BPH resistance in rice (*Oryza sativa* L.). Theor Appl Genet 112:288–297

Joshi SP, Ranjekar PK, Gupta VS (1999) Molecular markers in plant genome analysis. Curr Sci 77:230–240

Kalo P, Seres A, Taylor SA et al. (2004) Comparative mapping between *Medicago sativa* and *Pisum sativum*. Mol Genet Genomics 272:235–246

Kasai K, Morikawa Y, Sorri VA et al. (2000) Development of SCAR markers to the PVY resistance gene Ry_{adg} based on a common feature of plant disease resistance genes. Genome 43:1–8

Katiyar SK, Bennett J (2001) Biotechnology for gall midge resistance: from molecular tagging to gene pyramiding. In: Peng S, Hardy B (eds) Rice research for food security and poverty alleviation. IRRI, Manila, Philippines, pp 369–378

Kelly JD, Gepts P, Miklas PN et al. (2003) Tagging and mapping of genes and QTL and molecular marker-assisted selection for traits of economic importance in bean and cowpea. Field Crop Res 82:135–154

Khlestkina EK, Salina EA (2006) SNP markers: methods of analysis, ways of development and comparison on an example of common wheat. Russ J Genet 42:585–594

Koebner R, Summers R (2002) The impact of molecular markers on the wheat breeding paradigm. Cell Mol Biol Lett 7:695–702

Koebner R, Summers RW (2003) 21st century wheat breeding: plot selection or plate detection? Trends Biotechnol 21:59–63

Koebner R, Summers R (2007) Marker-assisted selection in wheat: evolution, not revolution. In: Guimarães E, Ruane J, Scherf B, Sonnino A, Dargie J (eds) Marker-assisted selection. Current status and future perspectives in crops, livestock, forestry and fish. Food and Agriculture Organization of the United Nations, Rome

Kota R, Varshney RK, Prasad M et al. (2007) EST-derived single nucleotide polymorphism markers for assembling genetic and physical maps of the barley genome. Funct Integr Genomics. doi: 10.1007/s10142–007–0060–9

Kwok PY, Chen X (2003) Detection of single nucleotide polymorphisms. Curr Issues Mol Biol 5:43–60

Labate JA, Baldo AM (2005) Tomato SNP discovery by EST mining and resequencing. Mol Breed, 16(4):343–349

Lagudah ES, Moullet O, Appels R (1997) Map-based cloning of a gene sequence encoding a nucleotide binding domain and a leucine-rich region at the *Cre3* nematode resistance locus of wheat. Genome 40:659–665

Landjeva S, Korzun V, Börner A (2007) Molecular markers: actual and potential contributions to wheat genome characterization and breeding. Euphytica 156:271–296

Lecomte L, Duffé P, Buret M et al. (2004) Marker-assisted introgression of five QTLs controlling fruit quality traits into three tomato lines revealed interactions between QTLs and genetic backgrounds. Theor Appl Genet 109(3):658–668

Liu J, Liu D, Tao W et al. (2000) Molecular marker-facilitated pyramiding of different genes for powdery mildew resistance in wheat. Plant Breed 119:21–24

Liu S, Anderson, JA (2003) Marker-assisted evaluation of *Fusarium* head blight resistant wheat germplasm. Crop Sci 43:760–766

Mammadov JA, Brooks WS, Griffey CA et al. (2007) Validating molecular markers for barley leaf rust resistance genes *Rph5* and *Rph7*. Plant Breed 126(5):458–463

Marczewski W, Flis B, Syller J et al. (2001) A major QTL for resistance to Potato leafroll virus (PLRV) is located in a resistance hotspot on potato chromosome XI and is tightly linked to N-gene-like markers. Mol Plant Microbe Interact 12:1420–1425

Melchinger AE, Utz HF, Schön CC (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics 149:383–403

Miklas PN, Kelly JD, Beebe SE et al. (2006) Common bean breeding for resistance against biotic and abiotic stresses: from classical to MAS breeding. Euphytica 147:105–131

Millán T, Clarke HJ, Siddique KHM et al. (2006) Chickpea molecular breeding: new tools and concepts. Euphytica 147:81–103

Mohan M, Nair S, Bentur JS et al. (1994) RFLP and RAPD mapping of the rice *Gm2* gene that confers resistance to biotype 1 of gall midge. Theor Appl Genet 87:782–788

Mohan M, Sathyanarayanan PV, Kumar A et al. (1997) Molecular mapping of a resistance-specific PCR based marker linked to a gall midge resistance gene (*Gm4t*) in rice. Theor Appl Genet 95:777–782

Nair S, Bentur JS, Rao UP et al. (1995) DNA markers tightly linked to a gall midge resistance gene (*Gm2*) are potentially useful for marker-aided selection in rice breeding. Theor Appl Genet 91:68–73

Oberhagemann P, Chatot-Balandras C, Schafer-Preg R et al. (1999) A genetic analysis of quantitative resistance to late blight in potato: towards marker-assisted selection. Mol Breed 5:399–415

Ordon F, Friedt W, Scheurer K et al. (2004) Molecular markers in breeding for virus resistance in barley. J Appl Genet 45(2):145–159

Paterson AH, Lander ES, Hewitt JD et al. (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721–726

Paterson AH, Tanksley SD, Sorrells ME (1991) DNA markers in plant improvement. Adv Agronomy 46:39–90

Pellio B, Streng S, Bauer E et al. (2005) High-resolution mapping of the *Rym4/Rym5* locus conferring resistance to the barley yellowmosaic virus complex (BaMMV, BaYMV, BaYMV-2) in barley (*Hordeum vulgare* ssp. vulgare L). Theor Appl Genet 110:283–293

Pernet A, Hoisington D, Franco J et al. (1999a) Genetic mapping of maize streak virus resistance from the Mascarene source. I. Resistance in line D211 and stability against different virus clones. Theor Appl Genet 99:524–539

Pernet A, Hoisington D, Dintinger J et al. (1999b) Genetic mapping of maize streak virus resistance from the Mascarene source. II. Resistance in line CIRAD390 and stability across germplasm. Theor Appl Genet 99:540–553

Picoult-Newberg L, Ideker TE, Pohl MG et al. (1999) Mining SNPs from EST databases. Genome Res 9:167–174

Prioul-Gervais S, Deniot G, Receveur E-M et al. (2007) Candidate genes for quantitative resistance to *Mycosphaerella pinodes* in pea (*Pisum sativum* L.). Theor Appl Genet 114(6):971–984

Rae SJ, Macaulay M, Ramsay L et al. (2007) Molecular barley breeding. Euphytica 158:295–303

Ragot M, Biasiolli M, Delbut MF et al. (1995) Marker-assisted backcrossing: a practical example. In: Bervillé A, Tersac M (eds) Les Colloques, No. 72, Techniques et utilisations des marqueurs moléculaires. INRA, Paris, France, 45–56

Ribaut JM, Jiang C, Hoisington D (2002) Simulation experiments on efficiencies of gene introgression by backcrossing. Crop Sci 42(2):557–565
Richardson KL, Vales MI, Kling JG et al. (2006) Pyramiding and dissecting disease resistance QTL to barley stripe rust. Theor Appl Genet 113:485–495
Rick, CM, Fobes J (1974) Association of an allozyme with nematode resistance. Tomato Genet Coop Rep 24:25
Robert VJM, West MAL, Inai S et al. (2001) Marker-assisted introgression of blackmold resistance QTL alleles from wild *Lycopersicon cheesmanii* to cultivated tomato (*L. esculentum*) and evaluation of QTL phenotypic effects. Mol Breed 8(3):217–233
Sanchez AC, Brar DS, Huang N, Li Z et al. (2000) Sequence tagged site marker-assisted selection for three bacterial blight resistance genes in rice. Crop Sci 40:792–797
Sardesai N, Kumar A, Rajyashri KD et al. (2002) Identification and mapping of an AFLP marker linked to *Gm7*, a gall midge resistance gene and its conversion to a SCAR marker for its utility in marker aided selection in rice. Theor Appl Genet 105:691–698
Schneider K, Kulosa D, Soerensen TR et al. (2007) Analysis of DNA polymorphisms in sugar beet (*Beta vulgaris* L.) and development of an SNP-based map of expressed genes. Theor Appl Genet 115(5):601–15
Servin B, Martin OC, M'ezard M et al. (2004) Toward a theory of marker-assisted gene pyramiding. Genetics 168(1):513–523
Sharma PN, Torii A, Takumi S et al. (2004) Marker-assisted pyramiding of brown planthopper (*Nilaparvata lugens* Stal) resistance genes *Bph1* and *Bph2* on rice chromosome 12. Hereditas 140:61–69
Simcox KD, Bennetzen JL (1993) Clustering of disease resistance loci in the maize genome. Phytopathology 83:1326–1330
Singh S, Sindhu JS, Huang N et al. (2001) Pyramiding three bacterial blight resistance genes (*xa5, xa13*, and *Xa21*) using marker-assisted selection into indica rice cultivar PR106. Theor Appl Genet 102:1011–1015
Slikova S, Gregova E, Bartos P et al. (2003) Marker-assisted selection for leaf rust resistance in wheat by transfer of gene *Lr19*. Plant Protect Sci 39:13–17
Somers DJ, Thomas J, DePauw R et al. (2005) Assembling complex genotypes to resist *Fusarium* in wheat (Triticum aestivum L.). Theor Appl Genet 111:1623–1631
Stein N, Perovic D, Kumlehn J et al. (2005) The eukaryotic translation initiation factor 4E confers multiallelic recessive Bymovirus resistance in *Hordeum vulgare* (L.). Plant J 42:912–922
Tanksley SD, Young ND, Paterson AH et al. (1989) RFLP mapping in plant breeding: new tools for an old science. Bio/Technol 7:257–264
Tanksley SD, Ganal MW, Prince JP et al. (1992) High-density molecular linkage maps of the tomato and potato genomes. Genetics 132:1141–1160
Timmerman, GM, Frew TJ, Weeden NF et al. (1994) Linkage mapping of *er*, a recessive *Pisum sativum* gene for resistance to powdery mildew fungus (*Erysiphe pisi*). Theor Appl Genet 88:1050–1055
Timmerman-Vaughan GM, Frew TJ, Butler R et al. (2004) Validation of quantitative trait loci for Ascochyta blight resistance in pea (*Pisum sativum* L.), using populations from two crosses. Theor Appl Genet 109:1620–1631
Tiwari KR, Penner GA, Warkentin TD (1998) Identification of coupling and repulsion phase RAPD markers for powdery mildew resistance gene *er*-1 in pea. Genome 41:440–444
Toojinda T, Baird E, Booth AA et al. (1998) Introgression of quantitative trait loci (QTLs) determining stripe rust resistance in barley: an example of marker-assisted line development. Theor Appl Genet 96:123–131
Torres AM, Román M, Avila CM et al. (2006) Faba bean breeding for resistance against biotic stresses: Towards application of marker technology, Euphytica 147:67–80
Tuvesson S, Post L, Tmhlund R et al. (1998) Molecular breeding for the BaMMV/BaYMV resistance gene *Ym4*. Plant Breed 117:66–22
van Berloo R, Aalbers H, Werkman A et al. (2001) Resistance QTL confirmed through development of QTL-NILs for barley leaf rust resistance. Mol Breed 8:187–195

Walker D, Boerma HR, All J et al. (2002) Combining *cry1Ac* with QTL alleles from PI 229358 to improve soybean resistance to lepidopteran pests. Mol Breed 9:43–51
Walker DR, Narvel JM, Boerma HR et al. (2004) A QTL that enhances and broadens Bt insect resistance in soybean. Theor Appl Genet 109:1051–1957
Wang XY, Chen PD, Zhang SZ (2001) Pyramiding and marker-assisted selection for powdery mildew resistance genes in common wheat. Acta Genetica Sinica 28:640–646 (in Chinese; summary in English)
Werner K, Friedt W, Ordon F (2005) Strategies for pyramiding resistance genes against the barley yellow mosaic virus complex (BaMMV, BaYMV, BaYMV-2). Mol Breed 16:45–55
Wicker T, Zimmermann W, Perovic D et al. (2005) A detailed look at 7 million years of genome evolution in a 439 kb contiguous sequence at the barley *Hv-eIF4E* locus: recombination, rearrangements and repeats. Plant J 41:184–194
Willcox MC, Khairallah MM, Bergvinson D et al. (2002) Selection for resistance to Southwestern corn borer using marker assisted and conventional backcrossing. Crop Sci 42:1516–1528
William HM, Morris M, Warburton M et al. (2007a) Technical, economic and policy considerations on marker-assisted selection in crops: lessons from the experience at an international agricultural research centre. In: Guimarães E, Ruane J, Scherf B, Sonnino A, Dargie J (eds) Marker-assisted selection. Current status and future perspectives in crops, livestock, forestry and fish. Food and Agriculture Organization of the United Nations, Rome
William HM, Trethowan R, Crosby-Galvan EM (2007b) Wheat breeding assisted by markers: CIMMYT's experience. Euphytica 157:307–319
Williams KJ, Willsmore KL, Olson S et al. (2006) Mapping of a novel QTL for resistance to cereal cyst nematode in wheat. Theor Appl Genet 112(8):1480–1486
Xu XF, Mei HW, Luo LJ et al. (2002) RFLP-facilitated investigation of the quantitative resistance of rice to brown planthopper (*Nilaparvata lugens*). Theor Appl Genet 104:248–253
Yang W and Francis DM (2005) Marker-assisted selection for combining resistance to bacterial spot and bacterial speck in tomato. J Am Soc Hort Sci 130(5):716–721
Yencho GC, Bonierbale MW, Tingey WM et al. (1996) Molecular markers locate genes for resistance to the Colorado potato beetle, *Leptinotarsa decemlineata*, in hybrid *Solanum tuberosum* × *S. berthaultii potato* progenies. Entomol Exp Appl 81:141–154
Young ND (1999) A cautiously optimistic vision for marker assisted breeding. Mol Breed 5:505–510
Yousef GG, Juvik JA (2001) Comparison of phenotypic and marker-assisted selection for quantitative traits in sweet corn. Crop Sci 41:645–655
Yu K, Park SJ, Poysa V (2000) Marker-assisted selection of common bean for resistance to common bacterial blight: efficacy and economics. Plant Breed 119:411–415

Chapter 9
Molecular Markers Based Approaches for Drought Tolerance

Deepmala Sehgal and Rattan Yadav

Abstract Drought stress is one of the most serious yield-reducing stresses in agriculture. Understanding the genetic basis of traits that can practically contribute towards development of drought resistant varieties is of paramount importance. This review focuses on the utilities of the chosen morpho-physiological traits in determining drought tolerance and provides an overview of molecular markers and genomics approaches that are available to increase the efficiency of breeding of these traits in crop breeding programmes. Suitable examples are cited where marker assisted selection (MAS) methods have started to prove useful in breeding for increased drought tolerance. Applications of model and non-model genomes in identifying functional markers associated with quantitative trait loci (QTLs), genes, and their allelic variants contributing to drought tolerance are highlighted. Genetics, genomics and molecular biology methods are also discussed for fine mapping and cloning of major QTLs as well as of their applications in developing more robust and affordable tools and technologies.

9.1 Introduction

Of all the abiotic stresses that curtail crop productivity, drought is the most devastating one from an economic standpoint and the most difficult to breeders' efforts. The difficulty arises from the diverse strategies adopted by plants themselves to combat drought stress depending on the timing, severity and stage of crop growth. Further complexity is added by the presence of other abiotic stresses (e.g., heat, salinity) which amplify the adverse effects of water deficit.

D. Sehgal and R. Yadav (✉)
Institute of Biological, Environmental and Rural Sciences (IBERS), Aberystwyth University, Gogerddan, Aberystwyth, Ceredigion, SY23 3EB, UK
e-mail: rsy@aber.ac.uk

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_9, © Springer Science+Business Media B.V. 2010

Compounding the problem further are the many loci that show efficacy only in a subset of circumstances (Lebreton et al. 1995; Ribaut et al. 1996, 1997; Tuinistra et al. 1996; Nguyen et al. 2004).

Understanding the traits that contribute to drought resistance in crops and their genetic basis is, therefore, fundamental to enable breeders and molecular biologists to develop new varieties with more drought resistance characters. Compared to conventional approaches, molecular mapping and genomics approaches offer unprecedented opportunities for dissecting quantitative traits like drought tolerance into their single genetic determinants, the so-called quantitative trait loci (QTLs,) thus paving the way to marker-assisted selection (MAS) and, eventually, cloning of QTLs and their direct manipulation via genetic engineering. The increasing number of studies reporting QTLs for drought-related traits and yield in drought-stressed crops indicates a growing interest in this approach. Furthermore, new genomics platforms, sequencing and bioinformatics have all added new dimensions for deciphering and manipulating the genetic basis of drought tolerance. This paper gives an overview of various molecular approaches that have been taken to study genetic basis of drought tolerance traits in plants and how breeders might improve the trait from the adoption of molecular approaches. Before embarking upon the applicability of molecular approaches for drought tolerance improvement, a brief overview of the traits associated with drought tolerance is outlined.

9.2 Traits Associated with Drought Tolerance

Whole plant response to drought stress is a complex process conditioned by a number of component responses that both interact and differ in their individual responses to the intensity and duration of water deficits. This review does not aim to analyze the morpho-physiological traits associated with drought tolerance in detail as a number of reviews have discussed them to a great depth (Tuberosa and Salvi 2004; Richards 2006; Cattiveli et al. 2008; Collins et al. 2008). Here we have discussed briefly the utilities of the chosen target traits in determining drought tolerance and their eventual manipulation in practical breeding applications.

Because of the low heritability of yield itself, the breeders are always looking for secondary traits associated with drought resistance that are determined more simply or having high heritabilities. However, it is arduous to distinguish between those traits that will actually improve yields under drought and traits that have no effect on yield. The development of advanced molecular marker technology is enabling a better understanding of the relationship of such traits (and their components) by dissecting them into underlying Mendelian units, the so-called QTLs (Lebreton et al. 1995; Prioul et al. 1997; Quarrie et al. 1999). By comparing the coincidence of QTLs for specific traits and QTLs for yield under drought stress (or yield stability) in the mapping population, it is possible to test much more precisely than before whether a particular yield (Ribaut et al. 1997; Yadav et al. 2002, 2004) and (or) yield contributing morpho-physilogical trait (Yadav et al. 1997;

Courtois et al. 2000; Tuberosa and Salvi 2006) is of significance in improving yield under drought conditions. Since abscissic acid (ABA) has been shown to be involved in regulating stomatal conductance, osmotic adjustment and root conductivity, interests have been shown in measuring ABA contents in order to establish relationships with drought resistance (Austin et al. 1982; Dingkuhn et al. 1991; Sanguineti et al. 1999). In maize, for example, 16 of the QTL regions influencing Leaf-ABA (L-ABA) also harboured QTLs for stomatal conductance, drought sensitivity index, leaf temperature, leaf relative water content, anthesis silking interval, and grain yield. The analysis of effects of each QTL region on the investigated traits indicated that L-ABA represented an indicator of level of drought stress experienced by the plant at the time of sampling because an increase in L-ABA was most commonly associated with a decrease in grain yield (Sanguineti et al. 1999). Similarly, evidence from wheat suggests that varieties which respond to drought by producing high concentrations of L-ABA do use less water and produce higher yields under drought (Innes et al. 1984). Roots are the main organs for plant water uptake, and development of deep and extensive root system is one of the adaptive strategies of plants for drought avoidance. In rice, manipulating root characters seems to be promising as putative QTLs for plant yield under drought were coincident with QTLs for root traits (Champoux et al. 1995; Yadav et al. 1997; Babu et al. 2003). Leaf rolling reduces the water loss in addition to reducing the leaf area exposed to heat and light radiation. There is some evidence that enhanced ability to roll leaves confers a yield advantage under drought conditions (Singh and Mackill 1991; Courtois et al. 2000). However, most breeders consider the triggering of leaf rolling as an indication of a plant suffering and select against its early manifestation.

Physiologists have considered water uptake (WU), water-use efficiency (WUE), and harvest index (HI) as drivers of yield and the secondary traits associated with drivers as proxy genetic markers, for instance, carbon-isotope discrimination (CID) for WUE, canopy temperature (CT) for WU, and anthesis-silking interval (ASI) for HI in maize (Reynolds and Tuberosa 2008). CT measured on wheat recombinant inbred lines (RILs) was associated with 60% of variation in yield under different drought prone environments (Olivares-Villegas et al. 2007). Economic analysis has confirmed the value of CT as an indirect selection tool to increase breeding efficiency (Brennan et al. 2007). Similarly, CID trait has shown promising applications in several crops (Condon et al. 2004). Spike photosynthesis in cereals, associated with high WUE, plays a major role in grain-filling under drought (Tambussi et al. 2007). Other sub-cellular processes such as photo-protective mechanisms including antioxidant systems and regulation of water flow via aquaporins seem to be equally effective in increasing WUE by improving metabolic efficiency (Reynolds and Tuberosa 2008).

Maize shows genetic variation in the relative timing of male and female 'readiness', referred to as ASI, a trait exacerbated by stress (Ribaut et al. 1996). Building on ASI-improved germplasm and the concept of selection under well-managed stress environments, a CIMMYT-coordinated breeding program resulted in significant impacts across southern Africa (Banziger et al. 2006).

9.3 Marker-assisted Selection for Drought Tolerance

The complexity of the drought phenomenon and the number of putative traits of unknown value has hampered progress in improving drought resistance through conventional methods. Moreover, most potential traits are difficult to screen accurately in large numbers, and such screening is often not possible under field conditions. Also, these traits are often controlled by many genes. Both factors greatly reduce breeding efficiency. It is perhaps not surprising, therefore, that this approach has had only limited success to date.

With the ready availability now of molecular markers it is possible to construct closely-spaced genetic maps of a particular genome using an appropriate mapping population of plants (Paterson et al. 1991; Tanksley 1993; Kumar 1999; Sehgal et al. 2008b). The availability of complete genetic maps of various crop species offers considerable advantage in overcoming some of the aforementioned problems associated with breeding varieties with improved yield under drought stress. Restriction fragment length polymorphism (RFLP) and more recently polymerase chain reaction (PCR)-based methods (Kumar 1999; Gupta and Rustogi 2004; Sehgal et al. 2008a) producing large numbers of genetic markers such as AFLP (Amplified Fragment Length Polymorphism), SSR (Simple Sequence Repeats), RAPD (Random Amplified Polymorphic DNA), STS (Sequence Tagged Sites), SAMPL (Selective Amplification of Microsatellite Polymorphic Loci) etc. allow loci controlling traits related to drought resistance to be identified and mapped in the genome. In practice, this can be true even when a large number of genes contribute to the total variation of a particular trait because the technique dissects the genome into small sections allowing a high degree of genetic resolution. The result is the identification of molecular markers which breeders can use as selection criteria, in place of a complex physiological phenomenon such as rooting depth or osmotic adjustment. Table 9.1 lists some of the major QTLs identified for drought tolerance traits and the linked genetic markers.

QTL discovery may in some cases be possible through a direct mRNA profiling approach (Hazen and Kay 2003), such as in the identification of the so-called eQTLs (QTLs influencing the level of expression of a particular gene). In this case, the analysis of the level of gene expression performed on each progeny of a mapping population will identify eQTLs influencing the observed variability among progenies in mRNA level of the profiled genes. Circumstantial evidence regarding the importance of each open reading frame (ORF) in governing variability for yield under conditions of drought can be obtained by comparing the map position of QTLs for yield with the map position of the ORFs itself and the corresponding eQTLs (Tuberosa et al. 2007). This approach has not been applied to identify eQTLs for drought-related traits. The cost associated with the profiling of the large number of RNA samples required to identify eQTLs is still too high to conceive a more routine application of this approach. Instead, transcriptome profiling is better suited for studies involving a limited number of samples, for instance, near isogenic lines (NILs) and/or bulked samples. Transcriptome profiling offers the opportunity of

Table 9.1 Major QTLs identified in crops for drought tolerance

Species	Trait	QTL	Chromosome/ LG	Cross (population)	Marker/marker-interval	Marker(s) type	Reference
Rice	Basal root thickness	brt4.1	4	IRAT 109 × Yuefu/ Jindao 187(NIL)	RM1136-RM273	SSR	Li Feng et al. (2007)
	1000-grain weight	tgw6.1	6		RM541-RM527		
	Grain yield	qgy4.1, 4.2 and 4.3	4	CT9993-510×X IR62266 (DH)	ME10_11-RZ565	RFLP,SSR, AFLP	Lancreas et al. (2004)
		qgy 3.1 and 3.2	3		EM11_9-RM231		
	Biological yield	qby4.1	4		RM273-RM317		
		qby8.1	8		ME2_11-G187		
	Harvest index	qhi3.1, 3.2, 3.3 and 3.4	3		RG104-RM81		
	Grain yield	qtl12.1	12	Vandana × Way Rarem (F3)	RM28048-RM511	SSR	Bernier et al. (2007)
	Osmotic adjustment	OA_{70}	8	Co39 × Morobereken (RIL)	RG1	RFLP	Lilley et al. (1996)
	Plant height	phs4.1	4	CT9993 × IR62266 (DH)	RG939-RG214	RFLP, AFLP, SSR	Babu et al. (2003)
	Grain height	gys4.1	4				
	No. of grains per panicle	gpps4.1	4				
Maize	ASI	QTL4	9	X178 × B73 (F_2)	Phio22	SSR	Li et al. (2004)
	Ear setting	QTL6	9		Umc 1657		
	Grain yield	QTL3	9		Umc1657		
Maize	Leaf ABA	L-ABA	2(bin2.04)	Os420 × IABO78 (F_4)	csu133	RFLP	Landi et al. (2005)
	Leaf ABA	L-ABA	2(1.03)	Os420 × IABO78 (F_4)	csu133	RFLP	Tuberosa et al. (1998)
Sorghum	Stay green	Stg2	LGA	B35 × TX7000 (RIL)	Cba-125H4_F	BAC derived markers	Sanchez et al. (2002)
	Stay green	Stg2	LGA	B35 × TX7000 (RIL)	W9889-R	RFLP	Xu et al. (2000)

(continued)

Table 9.1 (continued)

Species	Trait	QTL	Chromosome/ LG	Cross (population)	Marker/marker-interval	Marker(s) type	Reference
Barley	Plant height	PH	2H	Arta × *H. spontaneum* 41-1 (RIL)	HVM54-e38m49-11	AFLP, SSR	Baum et al. (2003)
			3H		e32m48–03-BMag0013		
			7H		E41m48–01		
Wheat	Grain yield	QYld.idw-2B[b]	2	Kofa × Svevo (RIL)	Xgwm1027-Xwmc361	SSR	Maccaferri et al. (2008)
		QYld.idw-3B[b]	3		Xbarc133-Xgwm493		
		QYld.idw-7B	7		Xgwm569-Xbarc1005		
	Heading date	QHd.idw-2A.2[b]	2		Xwmc177-Xcfa2201		
		QHd.idw-2B.2[c]	2		Xgwm1300-Xwmc332		
		QHd.idw-7B[b]	7		Xgwm569-Xbarc1005		
	Plant height	QPht.idw-1B.1[c]	1		Xbarc119-Xgwm413		
Wheat	Plant height	QPht.idw-2B[b]	2		Xgwm1027-Xwmc361	SSR	
		QPht.idw-3A[c]	3		Xgwm1159-Xgwm10		
		QPht.idw-7A[c]	7		Xcfa2028-Xbarc174		
Pennisetum	Grain yield	GY	2	H77/833-2 × PRLT2/89-33	Xpsm394-Xpsm214, Xpsm443-Xpsm356	RFLP	Yadav et al. (2002)
	Grain yield	GRYLD	2	ICMB841 × 863B	Xpsm458-Xpsmp2059	RFLP, SSR	Bidinger et al. (2007)

enriching the region with functional markers (FMs) designed from within the genes. FMs are superior to random markers such as RFLPs, SSRs and AFLPs owing to complete linkage with trait locus alleles and functional motifs (Andersen and Lübberstedt 2003). Once genetic effects have been assigned to functional sequence motifs, FMs derived from such motifs can be used to fix gene alleles (defined by one or several FM alleles) in several genetic backgrounds. This would be a major advance in marker applications, particularly in plant breeding, to select for example parent materials to build segregating populations, as well as for subsequent selections while constructing a variety (Andersen and Lübberstedt 2003). The availability of sequence data for genes through genome and (or) EST (Expressed Sequence Tag)-sequencing projects has made possible the development of FMs from transcribed regions of the genome (Gupta and Rustogi 2004). ESTs have been used to identify SSRs (Srinivas et al. 2008) and SNPs (Single Nucleotide Polymorphism; Rafalski 2002), and genic molecular markers have been developed in several species. In addition to being useful for identifying the 'perfect' or 'ideal' markers in MAS, FMs are important resources for estimating functional variation in natural or breeding populations (Gupta and Rustogi 2004; Sehgal and Raina 2008; Srinivas et al. 2008).

Molecular markers for trait selection have numerous advantages over morphological markers used in conventional plant breeding. MAS reduces or eliminates the reliance on specific environmental conditions during the selection phase, thus speeding up the process of conventional plant breeding and facilitating the improvement of traits which were difficult to improve by conventional methods. Based on the QTL mapping results, research efforts are underway aimed at MAS to validate the usefulness of this molecular breeding approach for drought tolerance improvement. Ribaut and Ragot (2007) used a marker assisted backcross in maize to introgress the favourable alleles at five QTLs explaining about 38% of the total phenotypic variance for the interval between anthers and silks extrusion (ASI) (Bolanos and Edmeades 1996), a trait negatively associated with yield in drought conditions. Grain yield of the best maize hybrids selected with molecular markers for four generations was, on an average, 50% higher than control hybrids under severe water stress conditions. The availability of molecular markers linked to the QTLs for ASI allows for a more effective selection under drought as well as when drought fails to occur at flowering (Ribaut et al. 2002, 2004). MAS has also been applied in maize to derive pairs of backcross derived lines differing for the parental alleles (Os420 and IABO78) at a major QTL (*root-ABA1*) on bin 2.04 near csu133 that was originally shown to consistently affect leaf ABA concentration (L-ABA; Tuberosa et al. 1998; Sanguineti et al. 1999) and, following its isogenization, also root architecture and other drought-related traits (Giuliani et al. 2005; Landi et al. 2005, 2007). A field evaluation conducted under well-watered and water-stressed conditions during two consecutive seasons indicated that each pair of *root-ABA1* backcross-derived lines differed significantly and markedly for L-ABA, thus confirming the effectiveness of MAS (Landi et al. 2005).

Similarly, when a pearl millet major QTL for grain yield under terminal drought stress was transferred into a drought sensitive genotype, the genotypes carrying the

introgression at the target QTL showed a consistent grain yield advantage (Serraj et al. 2005). Marker-assisted backcross (MABC) programs have been used to improve root traits affecting drought stress tolerance in the elite rice cultivars IR64 and Kalinga III (Shen et al. 2001; Steele et al. 2006). Foreground selection of the Azucena (an upland tropical japonica variety) allele at four QTLs for deeper roots was performed strictly on the basis of the genotypes at the marker loci up to the BC3F2. Selected NILs for the four target QTLs showed, depending of the target QTLs, improved root length (by 12% to 27% with respect to IR64) or improved deep root weight (two NILs had the highest phenotypic gain, outperforming IR64 by up to 75%; Shen et al. 2001). Five segments on different chromosomes were targeted for introgression into Kalinga III; four segments carried QTLs for improved root morphological traits (root length and thickness) and the fifth carried a recessive QTL for aroma (Steele et al. 2006). The target segment on chromosome 9 (RM242-RM201) significantly increased root length under both irrigated and drought stress treatments.

Stay-green is an important form of drought resistance mechanism in sorghum, conferring resistance to premature senescence under soil moisture stress during the post-flowering period. QTL studies have identified several genomic regions associated with resistance to pre-flowering and post-flowering drought stress (Haussmann et al. 2002; Sanchez et al. 2002). Three stay green QTLs, Stg1, Stg2 and Stg3, accounting for 20%, 30% and 16% of phenotypic variance, respectively, have been reported (Sanchez et al. 2002). Stg2, considered as the most important QTL, was found to be consistent across several different mapping populations and environments. For this QTL, near isogenic lines (NILs) have been developed by marker-assisted backcross breeding. Similarly, in rice NILs for two major QTLs, basal root thickness (BRT) and 1000-grain weight (TGW) were obtained by MAS through foreground selection for target QTL and background selection in three backcross generations (BC1F1, BC2F2 and BC3F3) (Li-Feng et al. 2007). This is an area of molecular marker technology which has started to show its impact towards the validation of major QTL identified for drought tolerance traits. Development of QTL-NILs is an important step for long term goals like QTL cloning and raising transgenics.

Comparative consensus maps, integrating information of anchor markers and the results of different mapping populations within the same species (Tuberosa et al. 2002; Sawkins et al. 2004; Pelleschi et al. 2006), provide useful information for MAS. In maize, chromosome bin 2.04 has been shown to influence a number of traits important for tolerance to drought in different backgrounds (Lebreton et al. 1995; Tuberosa et al. 1998; Quarrie et al. 1999; Sawkins et al. 2004; Landi et al. 2005). Bin 1.06 is another chromosome region that has consistently shown the presence of QTLs for grain yield and drought-related traits (e.g., root traits) in a number of genetic backgrounds (Lebreton et al. 1995; Tuberosa et al. 1998, 2003; Pelleschi et al. 2006). Similarly, a pearl millet LG 2 QTL for drought tolerance and yield in drought stress environments has been found consistent across different genetic backgrounds and is currently being pursued for MAS (Yadav et al. 2002, 2004;, Bidinger et al. 2007).

Recently, marker-assisted recurrent selection (MARS), a scheme based on successive generations of crossing individuals based on their molecular profile with the goal to attain a somehow ideal genotype at the different target QTL region, has refined MAS strategies (Peleman and van der Voort 2003). When compared to backcross-MAS, MARS allows for the selection of additional favourable alleles besides those targeted by backcross-MAS. For example, when the goal is to improve drought tolerance, MARS would enable the breeder to select for favourable alleles at yield QTLs in addition to the QTL alleles targeted for enhancing drought tolerance. This approach has successfully been used in maize breeding programs (Ragot et al. 2000; Johnson 2004; Crosbie et al. 2006) and is currently under development in pearl millet (Yadav et al. 2002, 2004; Bidinger et al. 2007).

A major factor that will greatly affect the extent to which MAS will be more routinely exploited in breeding programs relates to its cost-effectiveness as compared to conventional breeding practices. The application of high-throughput genotyping platforms based on the scoring of markers that do not require the use of gels (Salvi et al. 2001; Hardenbol et al. 2003; West et al. 2006) coupled with quick DNA extraction protocols are needed to streamline MAS and make it more cost-effective and widely applicable.

9.4 Candidate Genes for Drought Tolerance

Knowledge of the candidate genes underlying drought tolerance QTLs would be extremely useful both for the understanding of the biological basis of tolerance, and for utilization in MAS. Candidate genes (CGs) are defined as genes showing molecular polymorphisms and genetically associated with QTL or statistically associated with drought tolerance trait. The CGs can be divided into two broad groups based on their functions: genes involved in cell protection during drought stress and genes involved in the regulation of other genes involved in the drought responses. The first group includes proteins involved in for example osmotic adjustment, degradation, repairs, detoxification and structural adaptations. The second group contains regulatory proteins, such as protein kinases and transcription factors such as DREB, bZIP, MYB etc. Table 9.2 lists some of the important CGs for drought tolerance gathered from literature survey. The CG strategy shows promise to bridge the gap between quantitative and molecular genetics approaches to study drought tolerance.

The forward genetics approach for identifying underlying CGs derives from a known allelic difference conferring an improved phenotype. In such an approach, the objective is to identify a sequence change conferring the improved phenotype. Such a sequence change can then become the basis for a marker that is specific for that allele. These types of markers will always co-segregate with the trait of interest and should also be polymorphic in any cross. Such a marker will often be based on SNP. A complete set of markers that are based on SNPs or other sequence variations

Table 9.2 Candidate genes for drought stress tolerance

Candidate gene	Gene product	Function	Reference
DREB	Drought responsive element (DRE) binding factors	Transcription factor	Qin et al. (2004), Novillo et al. (2004), Oh et al. (2005) Ito et al. (2006)
OsCDPK7	Ca-dependent protein kinase	Signalling factor	Saijo et al. (2000)
ERA1	Farnesyl-transferase	Negative regulator of ABA sensing	Wang et al. (2005)
Mn-SOD	Mn-superoxide dismutase	ROS-scavenging proteins	McKersie et al. (1996)
AVP1	Vacuolar H+-pyrophosphatase	Ion transport	Gaxiola et al. (2001), Park et al. (2005)
HVA1	LEA proteins	Protective proteins	Bahieldin et al. (2005), Xiao et al. (2007), Chandra Babu et al. (2004)
OsLEA3	LEA proteins	Protective proteins	
ERECTA	A putative leucine rich repeat receptor kinase	Transcription factor	Masle et al. (2005)
otsA (otsB)	Trehalose-6-phosphate synthetase	Synthesis of osmolytes (Trehalose synthesis)	Garg et al. (2002)
P5CS	Pyrroline-5-carboxylate synthetase	Synthesis of osmolytes (Proline synthesis)	Zhu et al. (1998)
mtlD	Mannitol-1-phosphate dehydrogenase	Synthesis of osmolytes (Mannitol synthesis)	Abebe et al. (2003)
Adc	Arginic decarboxylase	Synthesis of osmolytes (Putrescine synthesis)	Capell et al. (1998)
ABF3 (ABF4)	ABRE (ABA-responsive element) binding factors	Transcription factor	Kang et al. (2002)
PARP	Poly ADP-ribose polymerase	ROS (reactive oxygen species)-scavenging mechanism	De Block et al. (2005)
MYB, MYC	MYB and MYC proteins	Transcription factors	Abe et al. (1997), Cominelli et al. (2005)
EXP15, EXP2, EXP13	Expansin proteins	Cell elongation proteins	Lee et al. (2001)
CIS	Cis acting element of gene pwsi 18 promotor	Regulatory factor	Joshee et al. (1998), Vinod et al. (2006)
RWC3	RWC	Water-channel proteins/ aquaporins	Vinod et al. (2006)
CRTDRE	C-repeat/drought-responsive element	Transcription factor	Vinod et al. (2006)

(continued)

Table 9.2 (continued)

Candidate gene	Gene product	Function	Reference
LTP	Lipid transfer protein	Membrane biogenesis	Kader J-C (1996), Vinod et al. (2006)
NADP-Me	NADP-malic enzyme	carbon metabolism	Laporte et al. (2002)
MAPKKK	Mitogen-activated protein kinase kinase kinase	Signalling factor	Shou et al. (2004)
AtNF-YB1	NF-YB proteins	Transcription factor	Nelson et al. (2007)
HRD	Hardy protein	Transcription factor	Karaba et al. (2007)
LFY	Leafy protein	Transcription factor	Tuyen and Prasad (2008)
Hv-WRKY38	WRKY protein	Transcription factor	Marè et al. (2004)
AREB1	bZIP (Basic domain leucine zipper)	Transcription factor	Fujita et al. (2005)
SRK2C	Protein kinase	Signalling factor	Umezawa et al. (2004)

could be developed. In the reverse genetics approach, SNPs are sought in candidate genes to identify the phenotypic effects of genes. Known SNPs can be used to identify new candidate genes through association mapping. Phenotypic differences that correspond to particular SNPs may be the result of the sequence change. These SNPs can then be used for MAS or screening germplasm and elite breeding lines. The nucleotide change that contributes to quantitative variation has been referred to as a quantitative trait nucleotide (QTN).

Recent progress in generation of high-density linkage maps saturated with a variety of molecular markers, genes and ESTs has led to a rapid progress in taking the CG approach. EST sequencing is the most efficient and cost-effective way of tagging most of the candidate genes long before complete genomic sequence is available. This approach has the additional appeal of focusing only on the most conserved portions of diverse genomes and on sequences that were likely to be of functional significance. After genetic mapping is accomplished, ESTs mapping very close to the trait QTLs are targeted for candidate gene analysis. A comparative analysis of QTL data for root traits of three rice mapping populations with one parent ('Azucena') in common, combined with the screening of ESTs and cDNA clones has identified two genes for cell expansion (OsEXP2 and EGase) that were mapped within the support intervals of QTLs for root traits common to the three populations (Zheng et al. 2003).

For those ESTs that could not be mapped because of a lack of polymorphism, the rice physical map will help to confirm their location. Markers can be designed from such ESTs for comparative and translational genomics by aligning the ESTs to rice BAC (Bacterial Artificial Chromosome) sequences and selecting candidate ESTs for designing conserved intron-spanning primers (CISPs; Feltus et al. 2006) This approach has provided easy to use, cost effective polymorphic markers to target a subset of candidate genes for co-segregation analysis. This approach is particularly useful for orphan crops for which very little genomic data or no DNA markers are available (Fredslund et al. 2006; Lohithasawa et al. 2007).

The identification of CGs for the drought tolerance QTLs can be facilitated by combining QTL maps with the functional maps enriched with genes potentially involved in controlling the target trait or with fully annotated genomic sequences. Sixteen candidate genes, out of 175 predicted, for drought tolerance have been identified in rice by integration of QTL map with rice genome physical map (Wang Xu et al. 2005). Specific efforts toward enriching linkage maps with function-specific genes have been undertaken in maize to identify QTL candidates (Davis et al. 2006). Over 6,000 root-specific expressed sequence tags have been mapped in silico on the consensus map (Davis et al. 2006). The routine integration of newly characterized genes on to a molecular map of the species from which the gene was isolated will facilitate the candidate gene approach as the functions of more genes are elucidated (Chao et al. 1994). Furthermore, advantage can be taken of extensive colinearity of blocks of genes (i.e. genes occurring in the same order) along chromosomes of related species. Thus the location of a structural gene for a particular enzyme in one species can be predicted on a related species, by comparative QTL mapping. For example, extensive colinearity has been identified amongst the cereal and grass genomes (Ahn and Tanksley 1993; Gale and Devos 1998). Osmotic adjustment (OA) of wheat was found to be influenced by alternate alleles at a single locus on chromosome 7A, with high response being recessive (Morgan and Tan 1996). Control of OA by the 7A locus was based primarily on potassium accumulation, and secondarily on amino acid accumulation. In rice, a QTL for OA under drought stress was identified on chromosome 8 across rice populations (Lilley et al. 1996; Robin et al. 2003). Comparative mapping indicates that the region of rice chromosome 8 containing the OA QTL is homeologous with a segment of wheat chromosome 7S which contains the OA locus identified by Morgan and Tan (1996) and with a barley chromosome 1 region where a QTL for relative water content in stressed conditions was identified (Teulat et al. 2003). Similarly, a rice QTL for OA located on chromosome 3 reside in a genomic region syntenic with the homeologous region of maize chromosome 1; in maize this region is associated with various physiological and agronomic traits affecting drought tolerance (Zhang et al. 2001). These results suggest that during cereal evolution, genes in these genomic regions in rice, wheat, barley and maize have been conserved to respond to drought conditions and might therefore contains useful candidate genes for the improvement of drought resistance in cereal crops.

Francki et al. (2004) and Li et al. (2004) used combinations of searches at the nucleotide and polypeptide levels (BLASTN, TBLASTX and BLASTX) in an attempt to account for UTRs (UnTranslated Regions), ORFs (Open Reading Frames) and diverged sequences that can be used as anchoring points for genome alignment. These authors carried out wheat-barley-rice comparisons based on publicly available genomic resources combining mapped wheat ESTs (Qi et al. 2003) and information from the first phase of the rice genome sequence (Goff et al. 2002; Yu et al. 2002). Stringent nucleotide and amino acid sequence alignments between wheat ESTs and rice sequences were used and the majority of orthologs were selected based on their highest sequence similarity.

The co-localization of specific genes with QTLs has proven to be an efficient approach to identifying the candidate genes for drought tolerance (Pelleschi et al. 1999). This approach has recently been used to identify six candidate genes for drought

tolerance in durum wheat (Diab et al. 2008). Osmotic potential at full turgor, transpiration, canopy temperature depression, water index, photosynthetic active radiation and grain carbon isotope discrimination have been proposed as selection criteria for drought in durum wheat (Rekika et al. 1998; Nachit et al. 2000). The genes M94726, AF519805 and D13042, coding for protein kinase, co-segregated with QTL for osmotic potential at full turgor and QTL for canopy temperature depression, quantum yield, chlorophyll content and water index in durum wheat. The loci Loxmjt and Lox11–1 coding for lipoxygenase, co-segregated with QTL for canopy temperature depression, photosynthetically active radiation, water index on chromosome 4B and with a QTL for photosynthetic active radiation on chromosome 5B, respectively. The locus AF210723 coding for fructan fructosyltransferase (FFT) is associated with a QTL for chlorophyll content on chromosome 6A (Diab et al. 2008). In barley, QTLs for osmotic potential and relative water content co-segregated with the Acl3 locus coding for the barley acyl carrier protein III on chromosome 1 (Teulat et al. 1998). This gene encodes a co-factor protein of the plant fatty acid synthetase involved in the *de novo* synthesis of the fatty acyl chain. It has a role in the protection of membranes during stress.

The vast amount of sequence information and/or mutants available for drought-related genes mapped/cloned in model and (or) non-model species provides additional opportunities for identifying candidate genes. This is particularly true for traits that have been more extensively investigated in species whose genome has been sequenced such as *Arabidopsis* and rice. A vast amount of information is available on the root transcriptome of *Arabidopsis* (Birnbaum et al. 2003; Birnbaum and Benfey 2004; Fizames et al. 2004). Although root development in *Arabidopsis* differs from other crop species in both overall architecture and the anatomy of individual roots, genes cloned in *Arabidopsis* could in some cases provide interesting leads for the identification of candidate genes for root QTLs, particularly for those functional and morphological features of root development that may have been conserved to a greater extent from an evolutionary standpoint. Recent work performed in barley (*Hordeum vulgare* L.) using a hairless mutant has led to the isolation and cloning of the β-expansin (EXPB) gene *HvEXP1* involved in root hair initiation (Kwasniewski and Szarejko 2006). Expansins are a large family of plant proteins endowed with a unique cell wall-loosening activity that have been shown to be involved in a number of processes related to plant growth and development (Li et al. 2003) which are likely to play an important role in cellular and organ elongation under different water availabilities. The redundancy of expansin-coding genes (58 in rice; Sampedro and Cosgrove 2005) points out the specificity of their role in growth and developmental processes, including root growth in maize (Wu et al. 1996, 2001). Based on these observations, a comparative analysis with the location of the QTLs reported for root hair length in maize (Zhu et al. 2005) with the map position of root-specific expansins will provide an opportunity to verify their role in controlling variability in root hair elongation.

Progress in the mass-scale profiling of the transcriptome, proteome and metabolome has allowed a more holistic approach in investigations of drought tolerance based on the measurement of the concerted expression of thousands of genes and their products. Seven thousand up- and down-regulated genes have been reported from gene expression analysis in *Arabidopsis* (Seki et al. 2002) under drought

stress that could serve as a resource for the identification of candidate genes for comparative genomic analysis of crop species. High-throughput techniques such as oligo chips, gene chips and various serial analysis of gene expression (SAGE) techniques are used for global gene expression analysis at a particular stage or time. The gridding of thousands of unique DNA sequences in large or small arrays provides a substrate that can be used to identify candidate genes that exert an influence at specific points in development. Common sources of DNA for the arrays include cDNA, ESTs, subgenomic regions of specific chromosomes, and even the entire set of genes in *Arabidopsis*.

9.5 Fine Mapping and Cloning of Drought Tolerance QTLs

An important step towards the application of molecular markers in breeding for drought tolerance is the cloning of DNA sequences underlying QTLs. It is viewed as an ideal entry point towards a more effective and informed exploitation of sequence variability at selected loci and to unlock the allelic richness present in germplasm collections, for instance, by means of EcoTILLING (Use of TILLING [Targeting Induced Local Lesions In Genomes] technique to survey natural variation in genes; Stemple 2004). A number of options are available to clone QTLs. To date, most plant QTLs have been cloned using a positional cloning approach following identification in experimental crosses. More recently, association mapping has received growing attention. Positional cloning and (or) map based cloning seeks an association between polymorphisms at marker loci and variability in the target quantitative trait while association mapping exploits linkage disequilibrium (LD) to identify the most promising candidate gene(s) for subsequent cloning.

Following the identification of a major QTL (Table 9.1), its positional cloning typically relies on the production of QTL-near isogenic line (QTL-NIL). The use of QTL-NILs, which isogenizes the QTL region, transformed the task of QTL cloning into one gene, similar to that performed for Mendelian traits. The major drawback of this approach is the long time required to obtain the NILs. Recently, a new method for identifying NILs has been validated in rice (Prabuddha et al. 2008). This method involves computation of simple correlation coefficients of all possible pairs of genotypes within a mapping population using molecular marker data, and phenotypic characterization of those pairs with very high positive correlation. The pairs showing very high correlation coefficients (0.70–0.97) and differing for less than 10% of the markers are considered as Genotypically Closely Related Pairs (GCRPs). Another important prerequisite for the positional cloning of a QTL is the availability of large number of DNA markers in the target region. The mapping resolution required for positional cloning is substantial because after primary mapping a QTL is positioned within a chromosome interval of ~10–30 cM, which usually includes several hundred genes. The recruitment of polymorphic markers required for fine mapping a QTL is fairly simple for *Arabidopsis* and rice because the whole genome has been sequenced, and also for species such as maize or tomato for which genomic

sequencing is under way or information is available in terms of, for example, ESTs and BAC ends. However, in species for which detailed sequence information is not available or cannot be deduced from syntenic relatives, many molecular markers that amplify multiple loci (e.g. AFLP, Random Amplification fingerprinting (RAF), SAMPL, DNA Amplification Fingerprinting (DAF) etc.) need to be screened in genotypes contrasted at the target region (e.g. pair of QTL-NILs).

Fine mapping can reduce the target locus in euchromatic regions to less than 100 kbp, a size that can be readily sequenced using standard BAC-based shotgun sequencing approaches. Shot gun sequencing of BAC-DNAs spanning the region followed by BLASTX analysis can be used to identify genes that are related to known protein coding genes. The BAC sequence can also be compared to the EST database to identify the transcribed portions of the BAC sequence. Alternatively, the low copy number DNA markers that are tightly linked to the gene of interest can be used as probes to screen large insert BAC libraries to identify appropriate BAC clones. Repeated rounds of such screening using low copy DNA markers from a series of BACs are required to identify overlapping clones by chromosome walking extending towards the target gene.

Reports of cloning major drought tolerance QTL locus are currently scanty despite high success rate of cloning QTLs/genes for other traits such as disease resistance (Yano et al. 2000; Kojima et al. 2002; Werner et al. 2005). A recent study has led to the cloning of major flowering-time QTL locus Vgt1 in maize through positional cloning and association mapping approach (Salvi et al. 2007). Efforts are also underway in maize to positionally clone *root-ABA1*, a QTL on bin 2.04 which affects root architecture and leaf ABA concentration (Giuliani et al. 2005; Landi et al. 2005). ERECTA gene, a sequence beyond a QTL for transpiration efficiency, has been cloned in *Arabidopsis* (Masle et al. 2005). The major hindrance to the cloning of major QTLs for most drought-related traits is due to their usually low to moderate heritability which requires an adequately large number of replicates to obtain accurate estimates of the phenotypic value of each segmental isoline.

The cloning of QTLs through association mapping relies on the molecular profiling and phenotypic characterization of natural populations. Unlike conventional bi-parental mapping populations which have been used for identifying QTLs for the trait of interest, the natural populations are the products of many cycles of recombination and have the potential to show enhanced level of resolution of QTLs. Strategies of association mapping are strongly influenced by the level of LD present in the target population. With high LD values, marker-trait association can theoretically be revealed with a manageable number of molecular markers. In this case, the expected mapping resolution will only be sufficient for the discovery and coarse mapping of the QTL. However, when testing germplasm panels with low LD, the diagnostic power of a single marker will only extend a short way and thus a prohibitively high number of markers would be required for a whole genome scan. Notably, the presence of population structure, namely the possible presence of hidden subgroups, due to selection and (or) relatedness, with an unequal distribution of alleles may influence the efficacy of this approach by causing spurious trait-marker associations (Pritchard et al. 2000).

Because SNPs offer the highest resolution for mapping QTL and are potentially in LD with the causative polymorphism they are the preferential candidate-gene variant to genotype in association studies (Rafalski 2002). Candidate-gene association mapping requires the identification of SNPs between lines and within specific genes. Therefore, the most straightforward method of identifying candidate gene SNPs relies on the re-sequencing of amplicons from several genetically distinct individuals of a larger association population. Fewer diverse individuals in the SNP discovery panel are needed to identify common SNPs, whereas many more are needed to identify rarer SNPs. Promoters, introns, exons, and 5′/3′-untranslated regions are all reasonable targets for identifying candidate gene SNPs, with non-coding regions expected to have higher levels of nucleotide diversity than coding regions (Zhu et al. 2008). It is not essential to score every candidate gene SNP. Because a key objective of this approach is to identify SNPs that are causal of phenotypic variation, those with a higher likelihood to alter protein function (coding SNPs) or gene expression (regulatory SNPs) should be a top priority for genotyping (Tabor et al. 2002). Completion of genome sequences and improved bioinformatics will facilitate in silico cross-matching of candidate sequences with QTLs in programmes of positional cloning or association mapping.

9.6 Concluding Remarks

The advent of molecular markers has enabled us to dissect quantitative traits into their single genetic components called QTLs, thus providing a more direct access to valuable genetic diversity of the morpho-physiological and agronomic traits contributing to yield under drought conditions. With the advances made in molecular markers, chances to evaluate trait-based approaches for addressing drought adaptation in crops appear to look wider than before.

More recently, bioinformatics and the deluge of information generated by sequencing and post-genomics platforms have added new dimensions for deciphering the role and function of genes governing the response to drought. Despite all of these impressive technological breakthroughs and the large number of QTLs shown to influence yield in drought-stressed crops, however, the overall impact of MAS and other genomics approaches on the release of drought-resilient cultivars has been marginal. This is partly due to our poor understanding of the molecular basis of drought tolerance and, most importantly, the difficulty in predicting the phenotypic value of a new genotype tailored through MAS. The success and effectiveness of MAS in assembling cultivars more tolerant to drought will rely on the identification of the relevant QTL alleles and their pyramiding in the correct combinations. Besides, the application of high-throughput genotyping platforms based on the scoring of markers that do not require the use of gels coupled with quick DNA extraction protocols are needed to streamline MAS and make it more cost-effective and widely applicable.

With the integration of QTL mapping, comparative mapping information, growing EST databases, expression results, and the identification of more and more genes the CG approach has become an important and powerful tool to uncover the mystery

behind the expression of quantitative traits for drought tolerance. Particularly, the comparative genetics proved to be extremely useful for the molecular dissection of drought tolerance QTLs in large, complex genome species (such as wheat and maize) by using co-linear regions from a smaller genome species (such as rice and sorghum). Through comparative mapping and bioinformatics it will hopefully be possible to utilize the vast knowledge emerging in model systems for the direct benefit of the closely related, lesser-studied ‘orphan’ crops. Extensive comparative genomics studies have already been completed amongst cereal species and much of this information is presented at the GRAMENE (http://www.gramene.org/) portal. Despite large difference in DNA content and chromosome number, the grass genomes maintain a high level of conserved macro-synteny and a moderate to high level of micro-synteny (Gale and Devos 1998). This has led to a diverse array of initiatives based on extrapolating and cross-referencing to rice as the model hub for the grass species. Recent advances in rice genomics research and completion of the rice genome sequence have made it possible to identify and map precisely a number of genes through linked DNA markers. Both macro-colinearity revealed by comparative mapping and micro-colinearity revealed by sequence comparisons indicate that sequencing and functional analysis of rice genome will have great impact for improvement of drought tolerance traits in other plants.

The identification of candidate genes is not the endpoint. Direct validation that a candidate gene causes variation in the trait under investigation is still required. Currently, transgenic plants have been used to test the effect of overexpression of specific gene known to be upregulated by drought stress. The incremental success of these experiments indicates a potentially useful role of these genes in achieving long term drought tolerance. Other technologies such as RNA interference (RNAi), EcoTILLING are now under development in order to accelerate candidate gene validation.

With its far-reaching implications, QTL cloning enables us to better understand and more effectively manipulate the traits influencing drought tolerance. It is foreseeable that QTL cloning will be facilitated by sequence information and the availability of new gene mapping techniques such as linkage disequilibrium/association mapping in future.

Acknowledgements The authors wish to acknowledge Biotechnology and Biological Sciences Research Council (BBSRC) and Department for International Development (DFID) for funding to their work via grant number BB/F004133/1

References

Abe H, Yamaguchi-Shinozaki K, Urao T et al. (1997) Role of Arabidopsis MYC and MYB homologs in drought- and Abscisic acid-regulated gene expression. Plant Cell 9:1859–1868

Abebe T, Guenzi AC, Martin B et al. (2003) Tolerance of mannitol-accumulating transgenic wheat to water stress and salinity Plant Physiol 131:1748–1755

Ahn S, Tanksley SD (1993) Comparative linkage maps of the rice and maize genomes Proc Natl Acad Sci U S A 90:7980–7984

Andersen JR, Lǜbberstedt T (2003) Functional markers in plants Trends Plant Sci 8:554–560
Austin RB, Henson IE, Quarrie SA (1982) Abscisic acid and drought resistance in wheat, millet and rice In: Drought resistance in crops with the emphasis on rice manila, IRRI, pp 171–180
Babu CR, Nguyen BD, Chamarerk V et al. (2003) Genetic analysis of drought resistance in rice by molecular markers: association between secondary traits and field performance Crop Sci 43:1457–1469
Bahieldin A, Hesham HT, Eissa HF et al. (2005) Field evaluation of transgenic wheat plants stably expressing the HVA1 gene for drought tolerance Physiol Plant 123:421–427
Banziger M, Setimela PS, Hodson D et al. (2006) Breeding for improved abiotic stress tolerance in maize adapted to southern Africa. Agric Water Manage 80:212–224
Baum M, Grando S, Backes G et al. (2003) QTLs for agronomic traits in the Mediterranean environment identified in recombinant inbred lines of the cross 'Arta' × *H spontaneum* 41–1 Theor Appl Genet 107:1215–1225
Bernier J, Kumar A, Ramaiah V et al. (2007) A large-effect QTL for grain yield under reproductive-stage drought stress in upland rice Crop Sci 47:507–518
Bidinger FR, Napolean T, Hash CT et al. (2007) Quantitative trait loci for grain yield in pearl millet under variable postflowering moisture conditions Crop Sci 47:969–980
Birnbaum K, Benfey PN (2004) Network building: Transcriptional circuits in the root Curr Opin Plant Biol 7:582–588
Birnbaum K, Shasha DE, Wang JY et al. (2003) A gene expression map of the Arabidopsis root. Science 302:1956–1960
Bolanos J, Edmeades GO (1996) The importance of the anthesis-silking interval in breeding for drought tolerance in tropical maize. Field Crop Res 48:65–80
Brennan JP, Condon AG, Van Ginkel M et al. (2007) An economic assessment of the use of physiological selection for stomatal aperture-related traits in the CIMMYT wheat breeding programme. J Agric Sci 145:187–194
Capell T, Escobar C, Liu H et al. (1998) Overexpression of the oat arginine decarboxylase cDNA in transgenic rice (*Oryza sativa* L) affects normal development patterns *in vitro* and results in putrescine accumulation in transgenic plants. Theor appl Genet 97:246–254
Cattivelli L, Rizza F, bedeck FW et al. (2008) Drought tolerance improvement in crop plants: an integrated view from breeding to genomics. Field Crops Res 105:1–14
Champoux MC, Wang G, Sarkarung S et al. (1995) Locating genes associated with root morphology and drought avoidance in rice via linkage to molecular markers. Theor Appl Genet 90:969–981
Chao S, Baysdorfer C, Heredia-Diaz O et al. (1994) RFLP mapping of partially sequenced leaf cDNA clones in maize. Theor Appl Genet 88:717–21
Chandra Babu R, Jingxian Z, Blumc A, David Hod T-H, Wue R, Nguyenf HT: HVA1, a LEA gene from barley confers dehydration tolerance in transgenic rice (Oryza sativa L.) via cell membrane protection. Plant Sci 2004, 166:855-862.
Collins NC, Tardieu F, Tuberosa R (2008) Quantitative trait loci and crop performance under abiotic stress: where do we stand. Plant Physiol 147:469–486
Cominelli E, Galbiati M, Vavasseur A et al. (2005) A guard cell specific MYB transcription factor regulates stomatal movements and plant drought tolerance. Curr Biol 15:1196–1200
Condon AG, Richards RA, Rebetzke GJ et al. (2004) Breeding for high water-use efficiency. J Exp Bot 55:2447–2460
Courtois B, McLaren G, Sinha PK et al. (2000) Mapping QTLs associated with drought avoidance in upland rice. Mol Breed 6:55–66
Crosbie, T.M., S.R. Eathington, G.R. Johnson, M. Edwards, R. Reiter, S. Stark, R.G. Mohanty, M. Oyervides, R.E. Buehler, A.K. Walker, R. Dobert, X. Delaney, J.C. Pershing, and M.A. Hall. 2006. Plant breeding: Past, present, and future. In K.R. Lamkey and M. Lee (ed.) Plant Breeding: Th e Arnel R. Hallauer Int. Symp., Mexico City. 17–22 Aug. 2003. Blackwell Publ., Ames, IA.
Davis D, Sanchez-Villeda H, Poroyko V et al. (2006) Root transcripts are functionally clustered in the maize genome In: Proceedings of the twenty-third annual missouri symposium plant

roots: from genes to form & function, 24–25 May 2006 University of Missouri, Columbia, MO, p 99
De Block M, Verduyn C, De Brouwer D, Cornelissen M: Poly(ADP-ribose) polymerase in plants affects energy homeostasis, cell death and stress tolerance. Plant J 2005,41:95-106.
Diab AA, Kantety RV, Ozturk NZ et al. (2008) Drought-inducible genes and expressed sequence tags associated with components of drought tolerance in durum wheat. Scientific Res Assay 3:9–26
Dingkuhn M, Cruz RT, O'Toole JC et al. (1991) Responses of seven diverse rice cultivars to water deficits III Accumulation of abscisic acid and proline in relation to leaf water-potential and osmotic adjustment. Field Crop Res 27:103–117
Feltus FA, Singh HP, Lohithaswa HC et al. (2006) A comparative genomics strategy for targeted discovery of single-nucleotide polymorphisms and conserved-noncoding sequences in orphan crops Plant Physiol 140:1183–1191
Fizames C, Mu S, Cazette C et al. (2004) The *Arabidopsis* root transcriptome by serial analysis of gene expression: Gene identification using the genome sequence. Plant Physiol 134:67–80
Francki MG, Appels R, Hunter A et al. (2004) Comparative organization of 3BS and 7AL using wheat-rice synteny. Funct Integr Genomics. doi 101007/s10142-004-0110-5
Fredslund J, Madsen LH, Hougaard BK et al. (2006) A general pipeline for the development of anchor markers for comparative genomics in plants. BMC Genomics 7:207
Fujita Y, Fujita M, Satoh R, Maruyama K, Parvez MM, Seki M, Hiratsu K, Ohme-Takagi M, Shinozaki K, Yamaguchi-Shinozaki K: AREB1 is a transcription activator of novel ABRE-dependent ABA-signaling that enhances drought stress tolerance in Arabidopsis, Plant Cell 2005, 17: 3470-3488.
Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci U S A 95:1971–1974
Garg A, Kim J, Owens T et al. (2002) Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses. Proc Natl Acad Sci U S A 99:15898–15903
Gaxiola RA, Li J, Undurraga S et al. (2001) Drought- and salt-tolerant plants result from overexpression of the AVP1 H+-pump. Proc Natl Acad Sci U S A 98:11444–11449
Giuliani S, Sanguineti MC, Tuberosa R et al. (2005) *Root-ABA1*, a major constitutive QTL, affects maize root architecture and leaf ABA concentration at different water regimes. J Exp Bot 56:3061–3070
Goff SA, Ricke D, Lan T-H et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L ssp *japonica*). Science 296:92–100
Gupta PK, Rustogi S (2004) Molecular markers from the transcribed/expressed region of the genome in higher plants. Func Integr Genomics 4:139–162
Hardenbol P, Baner J, Jain M et al. (2003) Multiplexed genotyping with sequence-tagged molecular inversion probes. Nat Biotechnol 21:673–678
Haussmann BIG, Mahalakshmi V, Reddy BVS et al. (2002) QTL mapping of stay-green in two sorghum recombinant inbred populations. Theor Appl Genet 106(1):133–142
Hazen SP, Kay SA (2003) Gene arrays are not just for measuring gene expression. Trends Plant Sci 8:413–416
Hsieh TH, Lee LT, Yang PT et al. (2002b) Heterology expression of the *Arabidopsis C-Repeat/ Dehydration Response Element Binding Factor 1* gene confers elevated tolerance to chilling and oxidative stresses in transgenic tomato. Plant Physiol 129:1086–1094
Innes P, Blackwell RD, Quarrie SA (1984) Some effects of genetic variation in drought-induced abscisic acid accumulation on the yield and water use of spring wheat. J agric Sci Camb 102:341–351
Ito Y, Katsura K, Maruyama K et al. (2006) Functional analysis of rice DREB1/CBF-type transcription factors involved in cold-responsive gene expression in transgenic rice. Plant Cell Physiol 47:1–13
Johnson, R. 2004. Marker-assisted selection. Plant Breed. Rev. 24:293–309.
Joshee N, Kisaka H, Kitagawa Y (1998) Isolation and characterization of a water stress-specific genomic gene, pwsi 18, from rice. Plant Cell Physiol 39:64–72

Kader J-C (1996) Lipid-transfer proteins in plants. Annu Rev Plant Physiol Plant Mol Biol 47:627–54

Kang JY, Choi HI, Im MY et al. (2002) *Arabidposis* basic leucine zipper proteins that mediate stress-responsive abscisic acid signalling. Plant Cell 14:343–357

Karaba A, Dixit S, Greco R et al. (2007) Improvement of water use effeiciency in rice by expression of HARDY, an *Arabidopsis* drought and salt tolerance gene. Proc Natl Acad Sci U S A 104:15270–15275

Kojima S, Takahashi Y, Kobayashi Y et al. (2002) Hd3a, a rice ortholog of the Arabidopsis FT gene, promotes transition to flowering downstream of Hd1 under short-day condition. Plant Cell Physiol 43:1096–1105

Kumar LS (1999) DNA markers in plant improvement: an overview. Biotech Adv 17:143–182

Kwasniewski M, Szarejko I (2006) Molecular cloning and characterization of β-expansin gene related to root hair formation in barley. Plant Physiol 141:1149–1158

Lancreas JC, Pantuwan G, Jongdee B et al. (2004) Quantitative trait loci associated with drought tolerance at reproductive stage in rice. Plant Physiol 135:384–399

Landi P, Sanguineti MC, Liu C et al. (2007) *Root-ABA1* QTL affects root lodging, grain yield and other agronomic traits in maize grown under well-watered and water-stressed conditions. J Exp Bot 58:319–326

Landi P, Sanguineti MC, Salvi S et al. (2005) Validation and characterization of a major QTL affecting leaf ABA concentration in maize. Mol Breed 15:291–303

Laporte MM, Shen B, Tarczynski MC (2002) Engineering for drought avoidance: expression of maize NADP-malic enzyme in tobacco results in altered stomatal function. J Exp Bot 53:699–705

Lebreton C, Lazic-Jancic V, Steed A et al. (1995) Identification of QTL for drought responses in maize and their use in testing causal relationships between traits. J Exp Bot 46 853–865

Lee Y, Choi D, Kende H (2001) Expansins: ever-expanding numbers and functions. Curr Opin Plant Biol 4:527–532

Li C, Ni P, Francki M et al. (2004) Genes controlling seed dormancy and pre-harvest sprouting in a rice-wheat-barley comparison. Funct Integr Genomics. doi 101007/s10142-004-0104-3

Li Y, Jones L, McQueen-Mason S (2003) Expansins and cell growth. Curr Opin Plant Biol 6:603–610

Li-Feng L, Hong-Liang Z, Ping M et al. (2007) Construction and evaluation of near-isogenic lines for major QTLs of basal root thickness and 1000-grain weight in lowland and upland rice. Chin J Agric Biotechnol 4:199–205

Lilley JM, Ludlow MM, McCouch SR et al. (1996) Locating QTL for osmotic adjustment and deydration tolerance in rice. J Exp Bot 47:1427–1436

Lohithasawa HC, Feltus FA, Singh HP et al. (2007) Leveraging the rice genome sequence for monocot comparative and translational genomics. Theor Appl Genet 115:237–243

Maccaferri M, Sanguineti MC, Corneti S et al. (2008) Quantitative trait loci for grain yield and adaptation of durum wheat (*Triticum durum* Desf) across a wide range of water availability. Genetics 178:489–511

Marè C, Mazzucotelli E, Crosatti C et al. (2004) Hv-WRKY 38: a new transcription factor involved in cold- and drought-response in barley. Plant Mol Biol 55:399–416

Masle J, Gilmore SR, Farquhar GD (2005) The ERECTA gene regulates plant transpiration efficiency in Arabidopsis. Nature 436:866–870

McKersie BD, Bowley SR, Harjanto E (1996) Water-deficit tolerance and field performance of transgenic alfalfa overexpressing superoxide dismutase. Plant Physiol 111:1177–1181

Morgan JM, Tan MK (1996) Chromosomal location of a wheat osmoregulation gene using RFLP analysis. Aust J Plant Physiol 23:803–806

Nachit MM, Monneveux P, Araus JL et al. (2000) Relationship of dryland productivity and drought tolerance with some molecular markers for possible MAS in durum wheat (*Triticum turgidum* L var durum) Durum wheat improvement in the Mediterranean region. New challenges pp. 203–206

Nelson DE, Repetti PP, Adams TR et al. (2007) Plant nuclear factor Y (NY-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. Proc Natl Acad Sci U S A 104:16450–16455

Nguyen TT, Klueva N, Chamareck V et al. (2004) Saturation mapping of QTL regions and identification of putative candidate genes for drought tolerance in rice. Mol Genet Genomics 272:35–46

Novillo F, Alonso JM, Ecker JR, Salinas J: CBF2/DREB1C is a negative regulator of CBF1/DREB1B and CBF3/DREB1A expression and plays a central role in stress tolerance in Arabidopsis. Proc Natl Acad Sci USA 2004, 101:3985-3990.

Oh SJ, Song SI, Kim YS et al. (2005) Arabidopsis CBF3/DREB1A and ABF3 in transgenic rice increased tolerance to abiotic stress without stunting growth. Plant Physiol 138:341–351

Olivares-Villegas JJ, Reynolds MP, McDonald GK (2007) Droughtadaptive attributes in the Seri/Babax hexaploid wheat population. Funct Plant Biol 2007:34:189–203

Park S, Li J, Pittman JK et al. (2005) Up-regulation of a H+-pyrophosphatase (H+-PPase) as a strategy to engineer drought-resistant crop plants. Proc Natl Acad Sci U S A 102:18830–18835

Paterson AH, Tanksley SD, Sorrells ME (1991) DNA markers in crop improvement In: Sparks DL (ed) Advances in agronomy, Academic, New York, pp 39–90

Peleman JD, van der Voort JR (2003) Breeding by design. Trends Plant Sci 8:330–334

Pelleschi S, Guy S, Kim JY et al. (1999) Ivr2, a candidate gene for a QTL of vacuolar invertase activity in maize leaves gene-specific expression under water stress. Plant Mol Biol 39:373–380

Pelleschi S, Leonardi A, Rocher JP et al. (2006) Analysis of the relationships between growth, photosynthesis and carbohydrate metabolism using quantitative trait loci (QTLs) in young maize plants submitted to water deprivation. Mol Breed 17:21–39

Prabuddha HR, Manjunatha K, Venuprasad R et al. (2008) Identification of near-isogenic lines: an innovative approach, validated for root and shoot morphological characters in a mapping population of rice (*Oryza sativa* L.). Euphytica 160:357–368

Prioul J-L, Quarrie S, Causse M et al. (1997) Dissecting complex physiological functions through the use of molecular quantitative genetics. J Exp Bot 48:1151–1163

Pritchard JK, Stephens M, Rosenberg NA et al. (2000) Association mapping in structured populations. Am J Hum Genet 67:170–181

Qi L, Echalier B, Friebe B et al. (2003) Molecular characterization of a set of wheat deletion stocks for use in chromosome bin mapping of ESTs. Funct Integr Genomics 1–2:39–55

Qin F, Sakuma Y, Li J, Liu Q, Li YQ, Shinozaki K, Yamaguchi-Shinozaki K: Cloning and functional analysis of a novel DREB1/CBF transcription factor involved in cold-responsive gene expression in Zea mays L. Plant Cell Physiol 2004, 45:1042-1052.

Quarrie S, Lazic Jancic V, Kovacevic D et al. (1999) Bulk segregant analysis with molecular markers and its use for improving drought resistance in maize. J Exp Bot 50:1299–1306

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94

Ragot, M., G. Gay, J.P. Muller, and J. Durovray. 2000. Efficient selection for the adaptation to the environment through QTL mapping and manipulation in maize. p. 128–130. In J.M. Ribaut and D. Poland (ed.) Molecular approaches for the genetic improvement of cereals for stable production in water-limited environments. A Strategic Planning Workshop, El-Batan, Mexico. 21–25 June 1999. CIMMYT, El-Batan, Mexico.

Rekika D, Nachit MM, Araus JL et al. (1998) Effects of water deficit on photosynthetic capacity and osmotic adjustment in tetraploid wheats. Photosynthetica 129–138

Reynolds M, Tuberosa R (2008) Translational research impacting on crop productivity in drought-prone environements. Curr Opin Plant Biol 11:171–179

Ribaut JM, Bänziger M, Betran J et al. (2002) Use of molecular markers in plant breeding: drought tolerance improvement in tropical maize In: MS Kang (ed) Quantitative genetics, genomics, and plant breeding CABI Publishing, Wallingford, UK, p 85–99

Ribaut JM, Bänziger M, Setter T et al. (2004) Genetic dissection of drought tolerance in maize: A case study In: Nguyen N, Blum A (ed) Physiology and biotechnology integration for plant breeding, Marcel Dekker Inc, New York, pp 571–611

Ribaut JM, Hoisington DA, Deutsch DA et al. (1996) Identification of quantitative trait loci under drought conditions in tropical maize: I Flowering parameters and the anthesis-silking interval. Theor Appl Genet 92 905–914

Ribaut JM, Jiang C, González-de-León D et al. (1997) Identification of quantitative trait loci under drought conditions in tropical maize: II Yield components and marker-assisted selection strategies. Theor Appl Genet 94 887–896

Ribaut JM, Ragot M (2007) Marker-assisted selection to improve drought adaptation in maize: the backcross approach, perspectives, limitations, and alternatives. J Exp Bot 58:351–360

Richards RA (2006) Physiological traits used in the breeding of new cultivars for water-scarce environments. Agric Water Manage 80:197–211

Robin S, Pathan MS, Courtois B et al. (2003) Mapping osmotic adjustment in an advanced backcross inbred population of rice. Theor Appl Genet 107:1288–1296

Saijo Y, Hata S, Kyozuka J et al. (2000) Over-expression of a single Ca2+-dependent protein kinase confers both cold and salt/drought tolerance on rice plants. Plant J 23:319–327

Salvi S, Sponza G, Morgante M et al. (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proc Natl Sci U S A 104:11376–11381

Salvi S, Tuberosa R, Phillips RL (2001) Development of PCR based assays for allelic discrimination in maize by using the 5-nuclease procedure. Mol Breed 8:169–176

Sampedro J, Cosgrove DJ (2005) The expansin superfamily. Genome Biol 6:242

Sanchez AC, Subudhi PK, Rosenow DT et al. (2002) Mapping QTLs associated with drought resistance in sorghum (*Sorghum bicolor* L Moench). Plant Mol Biol 48:713–726

Sanguineti MC, Tuberosa R, Landi P et al. (1999) QTL analysis of droughtrelated traits and grain yield in relation to genetic variation for leaf abscisic acid concentration in field-grown maize. J Exp Bot 50:1289–1297

Sawkins MC, Farmer AD, Hoisington D et al. (2004) Comparative map and trait viewer (CMTV): An integrated bioinformatic tool to construct consensus maps and compare QTL and functional genomics data across genomes and experiments. Plant Mol Biol 56:465–480

Sehgal D, Bhat V, Raina SN (2008a) Advent of DNA markers to decipher genome sequence polymorphism In: Kirti PB (ed) Handbook of new technologies for genetic improvement of grain legumes, CRC, New York, pp 477–495

Sehgal D, Bhat V, Raina SN (2008b) Applicability of DNA markers for genome diagnostics of grain legumes In: Kirti PB (ed) Handbook of new technologies for genetic improvement of grain legumes, CRC, New York, pp 497–557

Sehgal D, Raina SN (2008c) DNA markers and germplasm resource diagnostics: new perspectives in crop improvement and conservation strategies In: Arya ID, Arya S (eds) Utilization of biotechnology in plant sciences Rastogi, Meerut, India, pp 39–54

Seki M, Narusaka M, Ishida J et al. (2002) Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. Plant J 31:279–292

Serraj R, Hash CT, Rivzi SMH et al. (2005) Recent advances in marker-assisted selection for drought tolerance in pearl millet. Plant Prod Sci 8:334–337

Shen L, Courtois B, McNally KL et al. (2001) Evaluation of near-isogenic lines of rice introgressed with QTLs for root depth through marker-aided selection. Theor Appl Genet 103:75–83

Shou H, Bordallo P, Wang K (2004) Expression of the Nicotiana protein kinase (NPK1) enhanced drought tolerance in transgenic maize. J Exp Bot 55:1013–1019

Singh BN, Mackill DJ (1991) Genetics of leaf rolling under drought stress In: Rice genetics II, IRRI, Manila, pp 159–166

Srinivas G, Satish K, Murali Mohan S et al. (2008) Development of genic-microsatellite markers for sorghum staygreen QTL using a comparative genomic approach with rice. Theor Appl Genet 117:283–296

Steele KA, Price AH, Shashidhar HE et al. (2006) Marker-assisted selection to introgress rice QTLs controlling root traits into an Indian upland rice variety. Theor Appl Genet 112:208–221

Stemple DL (2004) TILLING: A high-throughput harvest for functional genomics. Nat Rev Genet 5:145–150

Tabor HK, Risch NJ, Myers RM (2002) Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet 3:391–397

Tambussi EA, Bort J, Guiamet JJ et al. (2007) The photosynthetic role of ears in C3 cereals: metabolism, water use efficiency and contribution to grain yield. Crit Rev Plant Sci 26:1–16

Tanksley SD (1993) Mapping polygenes. Annu Rev Genet 27:205–234

Teulat B, This D, Khairallah M et al. (1998) Several QTLs involved in osmotic-adjustment trait variation in barley (*Hordeum vulgare* L). Theor Appl Genet 96:688–698.

Teulat B, Zoumarou-Wallis N, Rotter B et al. (2003) QTL for relative water content in field grown barleys and their stability across Mediterranean environments. Thoer Appl Genet 108:181–188

Tuberosa R, Gill BS, Quarrie SA (2002) Cereal genomics: ushering in a brave new world. Plant Mol Biol 48:445–449

Tuberosa R, Salvi S (2004) QTLs and genes for tolerance to abiotic stress in cereals In: Gupta PK Varshney RK (eds) Cereal genomics, Kluwer Academic, Netherlands, pp 253–315

Tuberosa R, Salvi S (2006) Genomics-based approaches to improve drought tolerance of crops. Trends Plant Sci 11:405–412

Tuberosa R, Salvi S, Sanguineti MC et al. (2003) Searching for QTLs controlling root traits in maize: a critical appraisal. Plant Soil 255:35–54

Tuberosa R, Sanguineti MC, Landi P et al. (1998) RFLP mapping of quantitative trait loci controlling abscisic acid concentration in leaves of drought-stressed maize (*Zea mays* L). Theor Appl Genet 97:744–755

Tuberosa R, Silvio S, Silvia G et al. (2007) Genome-wide approaches to investigate and improve maize response to drought. Crop Sci 47 (S3) S120-S141

Tuinstra MR, Grote EM, Goldsbrough PB et al. (1996) Identification of quantitative trait loci associated with preflowering drought tolerance in sorghum. Crop Sci 36:1337–1344

Tuyen DD, Prasad DT (2008) Evaluating difference of yield trait among rice genotypes (*Oryza sativa* L.) under low moisture condition using candidate gene markers. Omonrice 16:24–33

Umezawa T, Yoshida R, Maruyama K, Yamaguchi-Shinozaki K, Shinozaki K: SRK2C, a SNF1-related protein kinase 2, improves drought tolerance by controlling stress-responsive gene expression in Arabidopsis thaliana. Proc Natl Acad Sci USA 2004, 101:17306-17311.

Vinod MS, Shrama N, Manjunatha K et al. (2006) Candidate genes for drought tolerance and improved productivity in rice. J Biosci 31:69–74

Wang Xu-S, Zhu J, Mansueto L et al. (2005) Identification of candidate genes for drought stress tolerance in rice by the integration of a genetic map (QTL) map with the rice genome physical map. J Zhejiang Univ SCI 6B (5):382–388

Wang Y, Ying J, Kuzma M et al. (2005) Molecular tailoring of farnesylation for plant drought tolerance and yield protection. Plant J 43:413–424

Werner JD, Borevitz JO, Warthmann N et al. (2005) Quantitative trait locus mapping and DNA array hybridization identify an FLM deletion as a cause for natural flowering-time variation. Proc Natl Acad Sci U S A 102:2460–2465

West MA, van Leeuwen H, Kozik A et al. (2006) High-density haplotyping with microarray-based expression and single feature polymorphism markers in Arabidopsis. Genome Res 16:787–795

Wu Y, Sharp RE, Durachko DM et al. (1996) Growth maintenance of the maize primary root at low water potentials involves increases in cell-wall extension properties, expansin activity, and wall susceptibility to expansins. Plant Physiol 111:765–772

Wu Y, Thorne ET, Sharp RE et al. (2001) Modification of expansin transcript levels in the maize primary root at low water potentials. Plant Physiol 126:1471–1479

Xiao B, Huang Y, Tang N et al. (2007) Over-expression of a LEA gene in rice improves drought resistance under the field conditions. Theor Appl Genet 115:35–46

Xu W, Subudhi PK, Crasta OR et al. (2000) Molecular mapping of QTLs conferring stay-green in grain sorghum (*Sorghum bicolour* L Moench). Genome 43:461–469

Yadav R, Courtois B, Huang N et al. (1997) Mapping genes controlling root morphology and root distribution in a doubled-haploid population of rice. Theor Appl Genet 94:619–632

Yadav RS, Hash CT, Bidinger FR et al. (2002) Quantitative trait loci associated with traits determining grain and stover yield in pearl millet under terminal drought stress conditions. Theor Appl Genet 104:67–83

Yadav RS, Hash CT, Bidinger FR et al. (2004) Genomic regions associated with grain yield and aspects of post-flowering drought tolerance in pearl millet across stress environments and testers background. Euphytica 136:265–277

Yano M, Katayose Y, Ashikeri M et al. (2000) Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS. Plant Cell 12:2473–2484

Yu J, Hu S, Wang J et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L ssp *indica*). Science 296:79–91

Zhang J, Zheng HG, Aarti A et al. (2001) Locating genomic regions associated with components of drought resistance in rice: Comparative mapping within and across species. Theor Appl Genet 103:19–29

Zheng BS, Yang L, Zhang WP et al. (2003) Mapping QTLs and candidate genes for rice root traits under different water-supply conditions and comparative analysis across three populations. Theor Appl Genet 107:1505–1515

Zhu BC, Su J, Chan MC et al. (1998) Overexpression of a d-pyrroline-5-carboxylate synthetase gene and analysis of tolerance to water-stress and salt-stress in transgenic rice. Plant Sci 139:41–48

Zhu C, Gore M, Buckler ES et al. (2008) Status and prospects of association mapping in plants. Plant Genome 1:5–20

Zhu J, Kaeppler SM, Lynch JP (2005) Mapping of QTLs for lateral root branching and length in maize (*Zea mays* L) under differential phosphorus supply. Theor Appl Genet 111:688–695

Chapter 10
Molecular Markers for Characterizing and Conserving Crop Plant Germplasm

G. Barcaccia

Abstract Molecular markers have revolutionized and modernized our ability to characterize genetic variation and to rationalize genetic selection, being effective and reliable tools for the analysis of genome architectures and gene polymorphisms in crop plants. The area of plant genomics that has shown the greatest development with respect to the use of molecular marker technology is that of population genetics. All DNA polymorphism assays have proven to be powerful tools for characterizing and investigating germplasm resources, genetic variation and differentiation of populations, on the basis of gene diversity and gene flow estimates. In the last decade, RFLP and PCR-derived molecular markers have also been extensively applied in plant genetics and breeding for Mendelian gene tagging and QTL mapping. As a matter of fact, the number of loci for which DNA-based assays have been generated has increased dramatically, the majority using PCR as methodology platform. The information acquired is now being exploited to transfer different traits, including biotic stress resistances and improved quality traits, to important varieties by means of marker-assisted selection (MAS) programs. Although the potential for take-up is now much wider than in the past, the progress seems nevertheless to be slow, albeit measurable. The most important challenges in the near future are certainly the molecular characterization of germplasm collections for preserving them from genetic erosion and the identification of phenotypic variants potentially useful for breeding new varieties. Knowing the presence of useful traits, genes and alleles would help in making decisions on the multiplication of plant accessions and the maintenance of seed stocks. There are no doubts that the use of molecular markers for characterization and conservation of genetic resources should be implemented so that potentially useful genes and genotypes can be added to core collections to make them exploitable by breeders. A new concept that might be successful is that of building crop plant collections primarily based on the knowledge of the presence of valuable genes and traits.

G. Barcaccia
Department of Environmental Agronomy and Crop Science, Faculty of Agriculture,
University of Padova, Campus of Agripolis, Viale dell'Università 16, 35020, Legnaro, (Italy)
e-mail: gianni.barcaccia@unipd.it

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_10,

10.1 Introduction

Molecular markers have proven to be powerful tools for analyzing germplasm resources and assessing genetic variation within as well as genetic differentiation among populations. In fact, the area of plant genomics that has shown the greatest development with respect to the use of DNA marker technology is that of population genetics. However, both RFLP and PCR-derived markers have also been extensively applied in plant genetics and breeding for mapping Mendelian genes and QTLs.

The use of molecular markers for investigating and managing genetic resources should be implemented so that useful information on genes and traits can be added to core collections to make them exploitable by breeders.

This chapter deals with the use of molecular markers for characterization of crop germplasm and for decision-making in conservation programs. Case studies related to important self- and cross-pollinated leguminous and cereal crops for exploring germplasm resources and mapping agronomic traits in landraces and elite stocks are presented. Theoretical considerations and experimental observations are critically discussed.

10.2 Genetic Characterization and Its Use in Decision-Making for the Conservation of Crop Germplasm: Basic Concepts

The characterization, conservation and exploitation of crop plant germplasm maintained in gene banks propound a number of challenges to the researchers dedicated to the investigation of plant genetic resources. Common problems include the development of strategies for sampling representative individuals in natural and experimental populations, the improvement of tools and technologies for long-term conservation and for high-throughput characterization of large numbers of stored accessions. The knowledge of the genetic diversity present in a gene bank is crucial for developing sustainable conservation strategies and it is also essential for the profitable exploitation of a gene bank by specific breeding programs. As a matter of fact, germplasm characterization of plant accessions deposited in gene banks has been limited and this likely represents a major cause for the limited adoption of conserved accessions in crop breeding programs (Ferreira 2006). Consequently, the genetic characterization of accessions belonging to a given collection and the examination of genetic relationships among them should be strengthened and perpetrated not only for maintaining but also for exploiting crop genetic resources.

Conservation of the genetic resources in the agro-ecosystem in which they have evolved (in situ conservation) is now being more widely considered, as complementary to strategies based on gene banks (ex situ conservation), for limiting genetic erosion and so preserving genetic diversity. If it is true that in situ conservation has been proposed essentially for wild relatives of cultivated plants,

it is also true that when considered for major crops this alternative can very often be unfeasible from a socio-economic perspective (Negri et al. 2000; Lucchin et al. 2003). Moreover, on-farm conservation of landraces is seen as a dynamic system that could help maintaining intact the genetic adaptation to changing conditions and the technical, social, cultural and environmental context in which they have occurred and evolved. In view of this, the whole of morpho-phenological and agronomic traits together with molecular markers could be the basis for the recognition of marks of specificity and typicality of landraces which would further enhance not only their on-farm conservation, but also allow their on-market valorization (Table 10.1).

Conservation of genetic resources entails several activities, many of which can greatly benefit from knowledge generated through the use of molecular marker technologies. The same applies to activities related to the acquisition and collection of germplasm (i.e., accurate description of accessions and application of effective procedures) and its evaluation for useful agronomic traits. The availability of a robust genetic characterization ensures that decisions made on conservation strategies will be supported by this information and results in an improved germplasm management. Of the experimental activities associated to plant genetic resources, the morpho-phenological and molecular evaluation of germplasm adds value to plant genetic resources and it is particularly important because helps the identification of genes and traits, and thus provides the basic knowledge for the exploitation of collections in programs aimed at breeding new crop varieties.

Classical attempts to directly use plant accessions stored in germplasm banks in breeding programs have been mainly focused to the identification of sources of genes of interest, such as resistance to plant pathogens or pests, and their transfer to cultivated materials. Linkage drag has very often restrained breeders from the initiative of using accessions from germplasm banks mainly because the improved material of advanced breeding programs is far more attractive than any germplasm resource of unknown genetic origin and phenotypic adaptation or performance (Ferreira 2006). When such a risk is taken, the accessions are usually screened to reveal the presence of a gene of interest and typically a backcross program is then initiated to transfer the gene to an elite line or cultivar. This procedure, however, is usually limited to Mendelian traits under simple genetic control, whereas complex traits require more elaborated methods, such as the mapping of QTLs in order to

Table 10.1 Main genetic, cultural and socio-economic features of on farm and ex situ conservation schemes for crop plant diversity

	Conservation strategy	
	On farm	Ex situ
Genetic drift and inbreeding	Operating	Not operating
Genetic adaptation to changing conditions	Happening	Not happening
Cultural and socio-economic role of crops	Maintained	Eroded
Cost	Moderate to high	Low to moderate

integrates the classical backcrossing program with linkage information based on the use of molecular markers.

The term "characterization" indicates the description of a single character or the evaluation of a set of genetic traits in individuals and populations. Moreover, this term is also used by means of distinguishing genotypes or gene pools. Thus, characterization of plant genetic resources refers to the process and tool by which accessions can be evaluated, differentiated and identified. In broad terms, this identification can refer to any morphological descriptor or molecular polymorphism of an accession. In the agreed terminology of gene banks and management of germplasm collections, characterization usually stands for the description of qualitative traits or quantitative traits that are highly heritable, easily scored by the eye and equally expressed in all environments (International Plant Genetic Resources Institute, IPGRI). Under a molecular point of view, characterization refers more specifically to the detection of DNA polymorphisms as a result of differences in random sequences or specific genes by using molecular marker techniques.

Standard characterization and evaluation of accessions can be routinely carried out by using different methods, including traditional practices such as the use of descriptor lists of morphological characters. They can also involve the adoption of agronomic trials under various environmental conditions. Genetic analysis and, in particular, molecular characterization of accessions refer to the visualization and description of DNA markers that follow Mendelian inheritance patterns and that involve specific or random sequences in the genome. In this context, the application of RFLP or PCR-derived markers for assaying polymorphisms at single or multiple loci all qualify as genetic characterization methods. Because of its nature, molecular characterization clearly offers an enhanced power for detecting diversity through fingerprinting and genotyping individual genomes and haplotyping individual genes. In addition, characterization with molecular markers offers a greater power of detection compared to phenotypic traits because they reveal differences at the genotypic level and are not influenced by the environment.

10.3 Use of Molecular Markers for the Characterization and Conservation of Plant Genetic Resources

Genomic DNA-based marker assays have revolutionized and modernized our ability to characterize genetic variation and to rationalize genetic selection (Lanteri and Barcaccia 2006). Molecular markers are known as particularly effective and reliable tools for the characterization of genome architectures and the investigation of gene polymorphisms in crop plants.

Besides linkage mapping, gene targeting and assisted breeding, the plant DNA polymorphism assays are powerful tools for characterizing and investigating germplasm resources and genetic relatedness. These techniques include restriction fragment length polymorphism (RFLP) markers and PCR-based molecular markers,

such as simple sequence repeat (SSR) or microsatellite markers (Morgante and Olivieri 1993), amplified fragment length polymorphism (AFLP) markers (Vos et al. 1995). SSR and AFLP markers are the most widely exploited techniques for the characterization of crop plant genetic resources: both types of markers offer an almost unlimited supply of molecular traits for distinctive fingerprinting and genotyping of plant materials, respectively, with and without a prior knowledge of target DNA sequences.

In general, molecular markers detect polymorphism by assaying subsets of the total amount of the DNA sequence variation in a genome. Polymorphisms detected by the RFLP assay reflect the variation of restriction fragment sites. PCR-based polymorphisms result from DNA sequence variation at primer binding sites and from DNA length differences between primer binding sites. The SSR assay utilizes pairs of primers flanking each simple sequence repeat and polymorphisms differ for the number of repetitive di-, tri- or tetranucleotide units present at one locus. The AFLP assay detects polymorphisms at multiple loci and involves the use of combination of primers specific for two distinct four-base and six-base long restriction sites flanking the target sequence unit. Although RFLP markers have provided useful estimates of the genetic diversity and relatedness in crop plants, there is some concern about their discriminatory power. Increasing the number of probe-enzyme combinations may improve the number of RFLP marker loci detectable, but the level of polymorphism that can be revealed by PCR-based markers still remains higher. In fact, owing to their own genetic nature, SSR markers usually detect multiple alleles at a given locus while AFLP assays mainly detect single alleles at multiple loci randomly distributed in the genome. A more recently introduced method is represented by SNP markers based on the detection of single-nucleotide polymorphisms by direct DNA sequencing of target gene regions (Fig. 10.1).

As a matter of fact, until now AFLP markers have provided the most widespread and robust technique with the highest polymorphism information content that combines the reliability of the RFLP technique with the potentiality of the PCR technique.

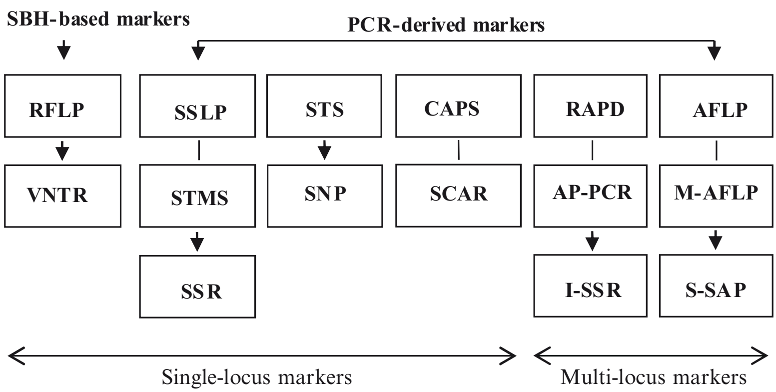


Fig. 10.1 Classification of the most commonly used molecular marker systems

Either AFLP and AFLP-derived markers, such as Microsatellite-AFLP and sequence-specific amplified polymorphism (S-SAP) markers, have been mostly used for fingerprinting and scanning whole genomes, characterizing single chromosomes, and tagging specific genes. However, microsatellite (SSR) markers along with SNP markers are now considered the most powerful and robust molecular marker systems for the analysis of whole genomes and single genes, and hence for the molecular characterization of single lines and population groups by means of genotyping or haplotyping.

Multi-locus marker systems, such as arbitrarily primed-PCR markers, AFLP and AFLP-derived markers visualize simultaneously many marker alleles: they supply an nearly unlimited number of polymorphisms and are exploitable over all species with no pre-existing genome or gene sequence knowledge, but show dominance (*i.e.*, only one allele identified, no possibility to discriminate between homozygous and heterozygous individuals). They exploit fingerprints, which are typically analyzed as pair-wise comparisons and whose results are to a limited extent reproducible and comparable among laboratories. Single-locus marker systems, such as RFLP, SSR and SNP markers, are usually characterized by co-dominance (*i.e.*, both alleles identified in heterozygous individuals) and thus are very informative in terms of polymorphisms, and supply reliable and comparable data. They allow to reconstruct genotypes and define haplotypes by multiplex analyses, however requires preliminary sequence information and thus are not always easily applicable.

Molecular markers are an irreplaceable tool to study biodiversity at the genetic level. Independently from the marker system considered, all types of plant DNA polymorphisms have proven to be powerful marker assays for assessing genetic variation and differentiation of populations, on the basis of gene diversity and gene flow estimates. In particular, the use of DNA markers allows to measure the genetic variation within single populations and to evaluate the genetic relatedness among populations, so that the formulation and implementation of germplasm maintenance and use programs can be optimized.

With the development of the PCR and DNA sequencing technologies associated with high-throughput screening systems, marker polymorphisms are now the choice for molecular-based surveys of genetic variation. Importantly, molecular markers showing different patterns of inheritance can now be investigated in nearly all of major crop species. The most widespread use of molecular markers in this context is the assessment of genetic variation within and genetic relationship among populations. Although in principle all types of molecular markers would be suitable for this purpose, microsatellite markers have been in the recent past and still are the most used in all the diversity studies. A standard set of markers for the major crop species should be selected and recommended to investigate the neutral genetic variability in the genome. In addition, one might also consider markers associated with important Mendelian genes and QTLs, reflecting the adaptive genetic potentials of individuals for a given qualitative or quantitative trait. For the most agriculturally important species, breeding research programs focus on the mapping of genes and polygenes so that an increasing number of markers for assisted characterization and selection of plant genetic resources will be available in the future.

Molecular markers are an indispensable tool to understand the genetic structures of populations. For the sampling of germplasm to create a plant gene bank, they are necessary but in no way sufficient to make adequate decisions. In addition to diversity information derived from molecular marker data sets, there are needs for understanding plant resource characteristics and specific knowledge on breeding values. It is therefore strongly recommended to concentrate co-ordinated phenotyping and genotyping efforts to fill the global maps of crop species diversity and to develop a better understanding of the rational decision-making process.

Molecular polymorphisms linked to quantitative traits or qualitative genes as well as nucleotide variants of causative genes can find practical application in marker-assisted selection (MAS) breeding programs.

Detailed linkage maps of specific chromosomes or chromosome blocks have been developed by analyzing the segregation of selected molecular markers of maternal and paternal origin in experimental populations. Moreover, wide-genome scanning approaches provide a method for rapidly identifying molecular markers linked to a specific genetic trait and accurately locating genes in a saturated genetic map. For instance, the identification of AFLP markers associated with specific genes based on the use of markers randomly distributed in the genome coupled with the use of near isogenic lines (NIL) or bulked segregant analysis (BSA) was applied not only in diploids but also in remarkably complex auto- and allo-polyploids. Both strategies are based on pooling the DNAs from individuals sharing the same genetic background but showing extreme classes of a given trait, and then screening for differences between genetic classes using molecular markers. Molecular differential screening of plants with contrasting characteristics is still considered one of the most powerful tools for identifying, isolating and using the genes underlying the expression of Mendelian traits and QTLs (quantitative trait loci). Similarly, natural populations can be exploited for discovering molecular markers linked to agronomically important genes by means of association mapping through linkage disequilibrium (LD) analysis. Interest in the study of LD, *i.e*. non-random association of alleles, in crop plants has increased dramatically in recent years because of two main factors (reviewed by Rafalski and Morgante 2004). First, genomic technologies enables rapid identification of haplotypes at many Mendelian loci, either by DNA sequencing or by detecting SNP markers. Second, in the presence of significant LD, it can be possible to identify genetic regions that are associated with a particular trait of interest (e.g., disease susceptibility) by genome scanning of individuals from an existing population. By contrast, if LD declines rapidly around the causative gene, the identification of genetic factors responsible for the trait of interest is possible by screening a limited set of candidate genes. Individual SNP markers or SNP haplotypes within a candidate gene are systematically tested for association with the phenotype of interest. Whole-genome scan and candidate gene approach are fundamentally similar methodologies, and differ primarily in the scale at which the analysis is performed.

An important point to be considered when using molecular markers in genetic characterization studies is the nature of the genomic DNA polymorphisms, since molecular markers can assay either neutral or adaptive variation. The genomics revolution of the last ten years has improved our understanding of the genetic

make-up of living organisms (Vendramin and Morgante 2006). Together with the achievements represented by complete genomic sequences for an increasing number of species, high throughput and parallel approaches are available for the analysis of transcripts, proteins, chemically-induced and transgenic mutants. All this information facilitates the understanding of the function of genes in terms of their relationship to the phenotype. Despite its great relevance, such an understanding could be of little value to population and conservation genetics because it will not elucidate the relationship between genetic variation in gene sequences and phenotypic variation in traits, but rather only that between a given gene and a mutant phenotype (Vendramin and Morgante 2006). The relationships between the phenotypic variation of complex traits and the molecular polymorphism of genes can be studied on the basis of a genomic approach. Work in model plant species such as Arabidopsis and rice has started to unveil a large number of genes involved in the determination of traits of adaptive significance, such as phenology and environmental stress tolerance or resistance. This progress will finally allow conservation genetics to directly analyze variation in genes involved in adaptive processes rather than in neutral markers. However, neutral markers will remain important to make inferences about stochastic processes affecting natural population evolution.

It is known that most molecular marker systems exploited for population genetics target genomic regions which are selectively neutral, even though molecular marker technologies which target specific genes do exist. The neutrality of markers is suitable for most uses in germplasm conservation and management. However, when the interest of conservation lies specifically in the diversity of traits of agronomic importance, some questions remain on the representativeness of markers. In such cases, the markers able to detect functional diversity are more suitable for the characterization and management of germplasm collections.

As a matter of fact, technology is rapidly evolving in molecular marker systems, moving from anonymous markers towards markers associated to a specific gene and/or a known chromosome position. One of the roles of biotechnology is that of supplying low-cost and high-throughput molecular tools for developing markers and making decisions in the processes of characterization and conservation of agricultural genetic resources.

10.4 Genetic Diversity and Similarity Statistics for Characterizing Plant Germplasm at the Population Level

Genetic diversity and similarity measurements are very useful for describing the genetic structure of populations. The genetic structure of natural populations of a crop plant species is strongly influenced by the reproductive system of their individuals and the union types occurring within populations. Breeding schemes that can be adopted as well as variety types that can be constituted depend on the reproductive barriers and mating systems of plants.

Natural populations of species that reproduce by apomixis or that propagate vegetatively are polyclonal, being composed by several genetically distinct clones and usually dominated by a few well-adapted genotypes. Therefore, genetic variation within populations is distributed among clones and most populations are characterized by different levels of differentiation among genotypes.

Landraces of self-pollinated species (e.g., bean, lentil, wheat and barley) are composed of a mixture of pure lines, genetically related but reproductively independent each other. Thus, genetic as well as phenotypic variation is mainly detectable among lines due to the presence within natural populations of fixed genotypes mainly homozygous for different alleles. Spontaneous hybridization is however possible to some extent depending on the species, environmental factors and germplasm stocks. Cultivated varieties of selfing species are usually represented by pure lines obtained by repeated self-pollination of a number of hybrid individuals stemmed from two parental lines chosen for complementary morphological and commercial traits.

Maize is one of the most commercially important cross-pollinated species. In many countries, existing landraces are selected by farmers for their own use and eventually sale to neighbors. Traditionally, landraces are developed by mass selection in order to obtain relatively uniform populations characterized by valuable production locally. Synthetics are also produced by intercrossing a number of phenotypically superior plants, selected on the basis of morpho-phenological and commercial traits. More rarely, plants are also evaluated genotypically by means of progeny tests. Compared to landraces, synthetics have a narrower genetic base but are equivalently represented by a heterogeneous mixture of highly heterozygous genotypes sharing a common gene pool. However, newly released varieties are exclusively represented by F_1 hybrids developed by private breeders and seed companies using inbred lines belonging to distinct heterotic groups.

Genetic characterization is providing new information to guide and prioritize conservation decisions for crop plants. The most urgently required action is the effective protection of all remaining wild ancestral populations and closely related species of crop plants, most of them now endangered. They are the only remaining sources of putative alleles of economic values that might have been lost during domestication events. It is equally important to ensure that the plant genetic resources selected for conservation include populations from the geographic areas representing the different domestication centres where high estimates of genetic diversity within and differentiation among populations are expected.

The assessment of the optimal plant and molecular marker sample size is a key step for the characterization of populations by computing genetic diversity and/or similarity statistics (Table 10.2). In particular, the choice of the most appropriate system (i.e., dominant vs codominant) and type (i.e., neutral vs functional) of markers and of the optimal number of markers and plants required to describe the genetic structure of a given population has to be carefully addressed.

For conservation purposes, individual genotypes of collected populations need to be identified to become part of the conservation scheme. Some general criteria can be defined concerning the desirable genetic properties of the sample: (1) it should represent the largest gene pool possible of the species; (2) it should

Table 10.2 Sizes of the plant sample for landraces in relation to the reproductive system of the species

Prevalent reproductive system of the species	Level and type of genetic variation in the population	Range of minimum sample size (seeds or plants)
Apomixis	Low (clones)	20–30
Amphimixis		
Selfing	Intermediate (pure lines)	40–60
Crossing	High (HW equilibrium)	80–100

have a maximum effective population size; (3) special genetic traits should be conserved. It is worth noting that maximizing the effective population size suggests collecting extreme genotypes, which may not be representative for the population. Using parameters from population genetics, the group of plants chosen should have minimum inbreeding and minimum relatedness among each other.

Two criteria should be followed for genetic marker-based characterization and conservation decision-making: priority populations for conservation should be those with the largest within-populations diversity and/or should maximize the conservation of between-populations diversity. Both within and between population diversity parameters are classically measured using molecular markers. In both cases, soundly based priority decisions for conservation at the global level will require the availability of large datasets.

Standard genetic diversity (H) and genetic differentiation (D) statistics of Nei (1973) and the inbreeding (F) coefficients of Wright (1965) are widely used to summarize the genetic structure of populations and the distribution of genetic variation using codominant molecular markers. The average (n_o) and the effective (n_e) number of alleles per locus well as the observed (H_o) and the expected (H_e) heterozygosity are the most commonly calculated population genetic parameters for the characterization of the within-subpopulation diversity. Let p_i denote the frequency of the i^{th} marker allele at a given locus, the genetic diversity computed as $H=1 - p_i^2$ is equivalent to the expected heterozygosity. This parameter can be computed for single subpopulations and the population as a whole. A hierarchical analysis of variance with estimation of F-statistics is then usually performed in order to measure the deficiency or excess of heterozygosity within subpopulation (F_{IS}) and between subpopulations (F_{IT}) as well the fixation index (F_{ST}) of the total population (Nei 1978). In particular, F_{ST} measures the effect of total population subdivision in different subpopulations and it corresponds to the reduction of heterozygosity of subpopulations compared to the population as a whole (Nei 1978) (Fig. 10.2).

Genetic similarity (GS) estimates between populations are also very often computed using dominant molecular markers. These estimates are traditionally calculated in all possible pair-wise comparisons between individuals either within or between populations using different genetic similarity parameters. For instance, genetic similarity estimates between individuals can be calculated using the following formula of Dice (1945): $GS = 2M_{ij}/(2M_{ij} + M_i + M_j)$, where M_{ij} represents the number of shared markers scored between the pair of individual fingerprints (*i* and *j*) considered, M_i is the number of markers present in *i* but absent in *j* and M_j

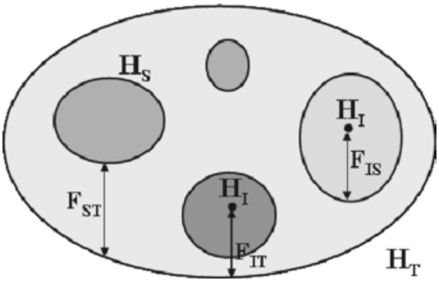


Fig. 10.2 Inbreeding (F) coefficients of Wright (1965) for assessing the genetic structure of populations

is the number of markers present in *j* but absent in *i*. Such estimate of genetic similarity corresponds to that calculated by the formula of Nei and Li (1979). The Simple Matching (SM) coefficient is also very popular for computing genetic similarity.

Any set of genetic diversities or similarities can be analyzed in terms of within and between population genetic variation, and more particularly, in terms of individual plant contributions to the total genetic variability. The most common approach used involves calculating a matrix of genetic distances or similarities on the basis of marker frequencies and marker fingerprint, respectively, generating UPGMA dendrograms and PCA centroids. Priority populations and individuals for conservation would therefore be the ones contributing most to the variation and differentiation of the germplasm set.

10.5 Using Molecular Marker-assisted Characterization and Conservation of Crop Plant Germplasm: Case Studies

Landraces are populations with high genetic variability and fitness to the natural and anthropological environments where they have originated. They represent not only a valuable source of useful traits, but also an irreplaceable bank of highly co-adapted genotypes. Knowledge of genetic variation within local populations and genetic differentiation with breeding stocks is expected to have a significant impact on the preservation and exploitation of plant germplasm resources. The case-studies reported in this paper refer to Italian maize and lentil landraces, durum and bread wheat pure lines, and Mesoamerican bean varieties.

In maize (*Zea mays* L.), unless in the last few decades conservation of local populations in gene-banks has taken place worldwide, the gradual replacement in cultivation of landraces by hybrids has resulted in genetic erosion. Many maize breeders are now concerned that genetic diversity within this species has been decreasing at an alarming rate as a consequence of modern hybrids and other agricultural changes. For instance, most inbreds have been developed from a limited number of elite lines and synthetics, a practice that heightens the risk of genetic uniformity in commercial maize production fields. Thus, maize breeders have

recently become more aware of the need for both assessing and maintaining genetic diversity among hybrid varieties and improving the management of genetic resources through the conservation of landraces. Information about genetic diversity and differentiation is known to be useful in planning crosses for hybrid development, assigning lines to some heterotic groups, maintaining genetic variability of landraces, and protecting inbreds and varieties. It can be obtained by surveying both qualitative and quantitative morphological traits or using molecular markers for investigating polymorphisms at the DNA sequence level.

Lentil (*Lens culinaris* Medik.) has been part of the human diet since a long time, being one of the first crops domesticated in the Near East. A great part of the landraces came from temperate areas of the Mediterranean basin, where lentils exist with colours that range from yellow to red-orange to green, brown and black. Two clear morphological groups based on seed characters could be established in this species, coinciding with the taxonomic description for the *macrosperma* and *microsperma* types. As in other Mediterranean countries, Italian landraces seem to be migrated following criteria related with seed characters preferred by farmers. Some of the landraces show an increasing importance in the food market of plant derivatives because lentil is one of the crops with the highest level of protein, being so a very important part of the diet.

Modern European wheat varieties are thought to display quite low levels of gene pool variation because of the high selective pressure applied in breeding programs. The genetic diversity of durum wheat (*Triticum turgidum* L. var. *durum*) and bread wheat (*Triticum aestivum* L. var. *aestivum*) elite germplasm has been traditionally estimated on the basis of morphological and quantitative traits, disease resistances, gliadin proteins and only recently by molecular markers. The adoption of molecular markers for the genetic characterization of wheat accessions is expected to play a key role not only for marker-assisted conservation but also for marker-assisted identification and selection of varieties in this species.

Regarding beans (*Phaseolus vulgaris* L.), yellow-coloured beans are among the traditional bean varieties grown principally in Mexico and Peru under several names, such as Azufrado and Canario. Originally, varieties from these countries represented two evolutionarily distinct bean groups as they originated from two different domestication centres, one in Mexico and the other in the southern Andes. Beans with yellow-coloured seeds are grown and consumed mainly in the north-western part of Mexico, but immigration from this country has created a market for yellow-seeded beans mainly in the USA and more recently also in Europe.

10.5.1 Genetic Anatomy of a Patented Yellow Bean (Phaseolus Vulgaris L.) Variety

Molecular markers find application in plant science and crop improvement to overcome limitations due to the absence of an appropriate and uniform legal protection of varieties and germplasm resources. In some countries, like USA, it is

possible to obtain a utility patent for varieties of crop species and it is also possible to obtain a Plant Variety Protection (PVP) certificate. Some of these awards can be controversial because of the perceived lack of novelty of distinctive morphological traits and the varieties themselves.

Genomic DNA fingerprinting was exploited by Pallottini et al. (2004) for assessing the genetic anatomy of a patented yellow bean (*Phaseolus vulgaris* L.), specifically the variety Enola. In order to check its origin, an AFLP-based genomic DNA fingerprinting study was carried out using a representative sample of 56 domesticated common bean accessions, including a subsample of 24 varieties with yellow coloured seeds morphologically similar to those of Enola. AFLP markers were detected using several primer combinations which revealed a total of 133 polymorphic markers. Most yellow-seeded beans, including Enola, were clustered in a tightly knit subgroup of the Andean gene pool. Enola was most closely related to the pre-existing Mexican cultivars: a sample of 16 individuals of Enola displayed full genetic identity with yellow-seeded beans from Mexico, namely Mayocoba and Azufrado Peruano (Fig. 10.3). Probability calculations of matching the specific Enola DNA fingerprint showed that the most likely origin of Enola is by direct selection within pre-existing yellow-bean cultivars from Mexico (Pallottini et al. 2004).

In particular, the least probable scenario was the one in which the Enola marker combination resulted from a cross between Andean and Mesoamerican genotypes, regardless of their seed colour represented in our sample (probability of 1×10^{-18}). The scenario with the highest probability represented selection without hybridization within cultivar Azufrado Peruano 87 (probability of 3×10^{-1}).

The assessment of the genetic anatomy of Enola as well as its genetic identity with pre-existing Central America local varieties raise questions about the rationales for the award of a utility patent and a PVP certificate.

10.5.2 Genetic Variation and Differentiation of Landraces of Lentil (Lens Culinaris Var. Microsperma L.) and Maize (Zea Mays Var. Indurata L.)

The Common Agricultural Policy of the EU concerning the quality of agricultural products and the preservation of landraces have driven more attention towards the Protected Geographical Indication (PGI) and Protected Designation of Origin (PDO) marks. Lentil (*Lens culinaris* var. *microsperma* L.) is extensively cultivated in the Mediterranean basin where it includes different landraces and modern varieties.

Barcaccia et al. (1998) demonstrated that the set up of molecular marker-based reference systems is feasible for the precise identification of single farmer's populations and for the preservation of the landrace as a whole. The mean Dice's genetic similarity estimates among lentil entries was shown to be 93%, ranging from 91% to 100%, when populations were analyzed as bulked DNA samples of 12 plants each. Although multi-locus DNA fingerprints were identical among most but not

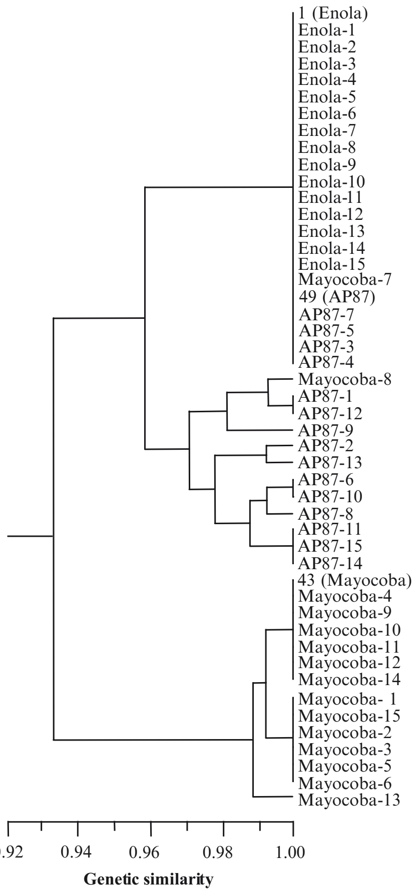


Fig. 10.3 UPGMA dendrogram showing the genetic relationships among Peruano-type bean cultivars: Enola, Azufrado Peruano 87 (AP87), and Mayocoba as assessed using AFLP markers (for additional information see Pallottini et al. 2004)

all the 26 chosen farmer's populations, indicating an homogeneous gene pool of the lentil landrace locally known as "Castelluccio of Norcia", some reliable polymorphic markers were visualized and three off-type farmer's populations were finally discovered. A number of polymorphisms were unambiguously shared with two Canadian (Laird) and Turkish (Kislik) commercial varieties used as genetic controls since their seeds were considered morphologically indistinguishable from the seeds of the Italian landrace under study. This finding cannot be explained by gene flow occurring among farmers but most likely by taking into account seed exchange with farmers who do not reproduce their own seed stocks. On the whole, the molecular approach proved to be essential for clearly identifying lentil local varieties and also for discriminating phenotypically similar imported lentil varieties.

Several types of molecular markers were also used for the characterization of the gene pool of Italian landraces of maize (*Zea mays* var. *indurata* L.). In particular, a comparative characterization of 10 field populations of the landrace "Nostrano of Storo" was carried out using different types of PCR-based markers (Barcaccia et al. 2003). The inbred line B73 and three synthetics (VA143,VA154 and VA157) selected from as many landraces were also used. Genetic diversity and relatedness were evaluated over 84 SSR and 53 Inter-SSR marker alleles using a total of 253 plants. Up to 23 alleles per SSR locus were scored while the average effective number of alleles per population was 6.99. Nei's total genetic diversity as assessed with SSR markers was $H_T = 0.851$ while the average diversity within populations was $H_S = 0.795$. The overall Wright's fixation index F_{ST} was as low as 0.066. Thus, more than 93% of the total variation was within population. Unique alleles over all SSR loci were found for six populations. An average of 17.7 marker alleles per Inter-SSR primer were scored with an effective number of marker alleles per locus of 1.34. The Shannon's diversity information index over all populations and I-SSR loci was 0.332, varying from 0.286 to 0.391. The extent of differentiation between populations was as low as $G_{ST} = 0.091$. Dice's genetic similarity matrices were estimated for both SSR and Inter-SSR markers. The mean genetic similarity coefficients within and between populations were respectively 0.269 and 0.217, for SSR markers, and 0.591 and 0.564, for Inter-SSR markers. UPGMA dendrograms displayed all field populations but one clustered into a distinct group, in which the synthetic VA154, selected from the "Marano Vicentino" landrace, was also included. One field population and the other two synthetics were clustered separately as well B73. The matrix correlation assayed by the Mantel's correspondence test was as high as 0.908. Findings suggest that, although a high variability can be found among plants, most plant genotypes belong to the same landrace locally called "Nostrano of Storo" (Barcaccia et al. 2003). Although gene flow from commercial hybrids might have occurred, the large number of polymorphisms and the presence of both unique alleles and alleles unshared with B73 and synthetics are the main factors underlying the value of this flint maize landrace as a source of genetic variation and peculiar germplasm traits. Because of its exclusive utilization for human consumption, such a molecular marker characterization will be a key step for obtaining a quality mark and so promoting the conservation and protection of the landrace.

The construction of a linkage map was then undertaken using several molecular marker systems, including AFLP and SSR markers (Barcaccia et al. 2006) (Fig. 10.4).

A set of mapped markers was adopted to study the effects of different conservation strategies (on farm, in situ and ex situ) on the population genetic structure of landraces (Pallottini 2002). Marker alleles showing significant frequency changes were distributed throughout the genome, but chromosome blocks carrying marker loci with differentiated genetic diversity values were detected. Some molecular markers revealed to be suitable also for the evaluation of the extent of gene flow that can occur between landraces and modern varieties.

The possibility of identifying the local populations of lentil and maize through their molecular characterization can be an essential element not only for conservation and breeding purposes, but also for certifying typical local products in order to

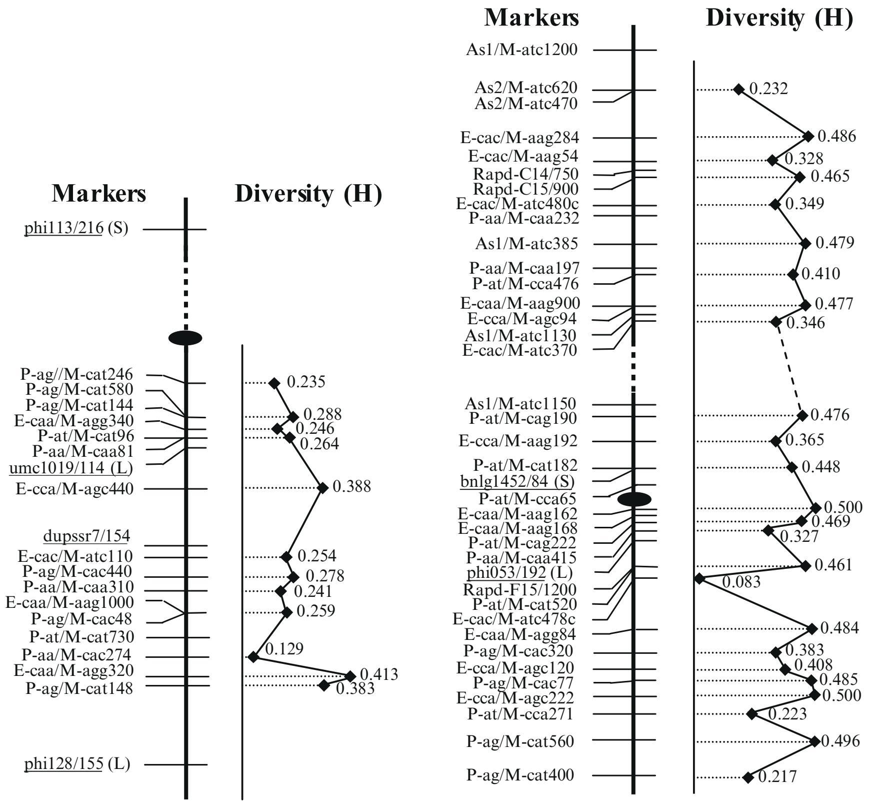


Fig. 10.4 Linkage groups of two maize chromosomes (5 and 3) carrying marker loci with differentiated genetic diversity (H) estimates (0.281 vs 0.409, on average)

avoid market frauds and to safeguard their gene pools. In the near future it could represent a basic requisite for their use in a serious and consumer-oriented production and marketing context.

10.5.3 Genetic Fingerprinting Durum Wheat (Triticum Durum l.) and Bread Wheat (Triticum Aestivum l.) Elite Germplasm Stocks for Multiple Breeding Purposes

The genetic diversity of durum wheat (*Triticum turgidum* L. var. *durum*) and bread wheat (*Triticum aestivum* L. var. *aestivum*) modern variety collections was estimated on the basis of molecular markers using AFLP technology for genomic DNA fingerprinting.

A total of 38 durum wheat and 26 bread wheat genomic DNA samples isolated from commercial varieties and experimental lines were investigated by fluorescent

AFLP markers using five different *Pst*I/*Mse*I and *Eco*RI/*Mse*I primer combinations previously selected on the basis of their ability of detecting polymorphisms. As many as 267 clearly detectable markers were scored, of which 59 (41.6%) and 73 (51.4%) proved to be polymorphic among varieties within and between species, respectively. Dice's genetic similarity (GS) estimates among the 64 pure lines were calculated in all possible pair-wise comparisons and the correspondent similarity matrix was used for the construction of UPGMA dendrograms and for the definition of centroids according to PCA analysis. Mean genetic similarity estimates within durum wheat and within bread wheat were 92% and 89%, respectively (Barcaccia et al. 2005). In each species, a few multi-locus genotypes showing almost full identity were found (Barcaccia et al. 2005). Several species-specific and variety-specific DNA markers were also scored: the latter types will be cloned, sequenced and converted into easily detectable single-locus markers.

On the whole, more than 68% of the total genetic variation found in wheat materials was explained by the first two principal coordinates. The observed number (n_o) and the effective number (n_e) of alleles were equal to 1.416 and 1.163 in durum wheat and to 1.514 and 1.184 in bread wheat, respectively. Nei's genetic diversity (H) estimates over all genomic loci were also comparable for the two species (0.102 and 0.119, respectively). Linkage disequilibrium (LD) tests were performed for all pair-wise comparisons of marker alleles. The number of significant LD was 78 over 142 loci (0.78%) in durum wheat and 139 over 142 loci (1.42%) in bread wheat. Preliminary data suggest the finding of a few AFLP markers displaying highly significant linkage disequilibrium ($P < 0.01$) with a number of wheat resistance genes, including yellow (stripe) and brown (leaf) rust, powdery mildew, Fusarium head blight and Septoria leaf spot diseases (Table 10.3).

Table 10.3 Estimates of linkage disequilibrium (LD) between molecular markers and resistance factors in bread and durum wheat varieties

Marker	Pathogen	LD	*r*	χ^2	*P*
A22	BR	−0.0831	−0.3508	4.68	0.0306
D51	YR	−0.0893	−0.3613	4.96	0.0259
A22	FUS	−0.1177	−0.5026	9.60	0.0019
D36	SEP	0.0693	0.3885	5.74	0.0166
E45	PM	−0.0436	−0.3254	4.02	0.0049
B27	BR	−0.0917	−0.0167	12.91	0.0003
D67	BR	−0.0651	0.0108	11.92	0.0006
C23	YR	0.0799	0.7826	7.63	0.0057
D15	YR	−0.0799	−0.4793	7.63	0.0057
C32	FUS	0.0888	0.8696	11.30	0.0008
E41	FUS	0.1154	0.6500	7.80	0.0057
C37	SEP	0.1154	0.6500	7.80	0.0052
D62	SEP	−0.1006	−0.3353	8.35	0.0052
C37	PM	0.1346	0.5948	8.33	0.0039
A06	PM	−0.1228	-0.0737	7.10	0.0077

YR: yellow (stripe) rust; BR: brown (leaf) rust; PM: powdery mildew; FUS: Fusarium head blight; SEP Septoria leaf spot.

This work is currently in progress. As future perspectives, additional *Pst*I/*Mse*I primer combinations will be assayed in the same set of wheat materials. It is known that *Eco*RI and *Pst*I enzymes sample different regions of the wheat genome: *Pst*I is methylation-sensitive and cuts principally in unmethylated DNA regions of the genome, containing expressed and mainly single-copy genes, whereas *Eco*RI is methylation-insensitive and cuts DNA throughout the genome. The final aim is that of assembling a database of DNA polymorphisms for the durum and bread wheat germplasm. This information is potentially useful for tracing single pure lines through genetic fingerprints and also for planning experimental crosses between pure lines on the basis of their genetic distances. Moreover, selected AFLP markers potentially linked to resistance genes will be converted into sequence-tagged site markers, such as sequence characterized amplified region (SCAR) or cleaved amplified polymorphic sequence (CAPS) markers, and then assayed in experimental populations segregating for the resistance traits in order to test their suitability and reliability for use in wheat MAS programs.

10.5.4 Effects of Different Conservation Strategies on the Population Genetic Structure of Maize Landraces as Assessed with Molecular Markers

The on farm, in situ and ex situ conservation methods may exert a different influence on the genetic structure of populations grown by farmers. This influence should be accurately evaluated to avoid genetic erosion and conservation programme failure. In fact, the loss of genetic diversity could be due to inbreeding that can result from drift and migration to natural and human selection and gene flow. Each of these factors has a different relative importance on the types of conservation methods. Molecular markers were used to investigate the influence of the conservation strategy on the genetic structure of farmer populations grown for 2 years with three different methods: (1) on farm conservation by farmers, using own seeds and traditional agronomic practices; (2) in situ conservation in the original area but taking into account the spatial isolation from other fields cultivated with hybrid varieties; and (3) ex situ conservation far from the original area with no gene flow due to the total absence of fields grown with the same crop (Pallottini 2002). Statistical tests failed to reveal any significant difference in terms of diversity/similarity absolute values among the populations conserved according to the three distinct strategies (Fig. 10.5).

Nevertheless, about 10% of the comparisons performed for the marker allele frequency parameter at the total assayed loci showed significant differences. Even the differences between genetic variation parameters computed for mapped and random marker loci were significant. In particular, some marker loci were more affected than others by changes of the marker allele frequency depending on the conservation method. These markers, distributed throughout the genome, could be related to important genes involved in the adaptation to environmental conditions or responsible for traits evaluated in the selection by farmers.

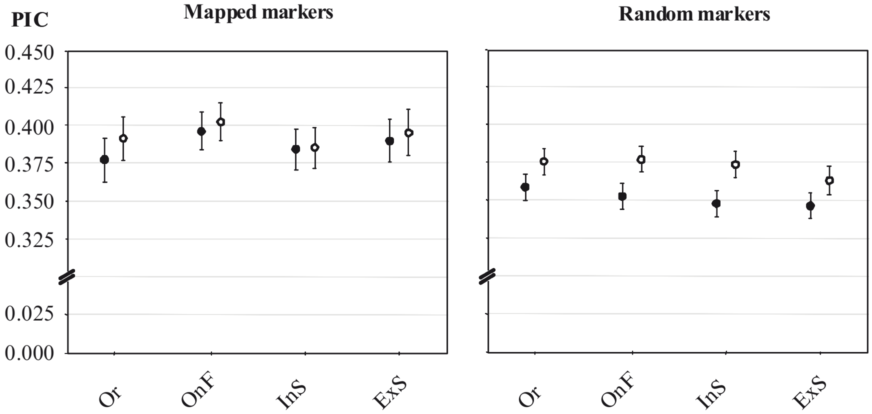


Fig. 10.5 PIC values and standard errors computed in the original population and in the populations obtained from on farm, in situ and ex situ conservation strategies, using two sets of mapped and random markers

In sum, although all conservation methods studied have determined the significant changes to the genetic structure of the farmer populations, the genetic variation and diversification that occurred with ex situ conservation was much stronger than that observed for in situ and on-farm conservation. It is worth mentioning that to monitor these changes, the level at which the investigation is performed is essential. When the mean values of the more common genetic diversity and/or similarity indexes are taken into account, no significant differences are highlighted because of the large set of molecular data and the occurrence of bidirectional changes of marker allele frequencies over all marker loci. Consequently, variation of the marker allele frequency has to be computed and interpreted at each single marker locus or between pairs of marker loci, but not on the whole molecular marker data set. We identified chromosome blocks in linkage groups 2, 3, 9 and 10 carrying marker loci affected by strong LD that could be associated to genes influenced by selection and that could have played a role in the adaptation to the different environmental situations (Barcaccia et al. unpublished results).

10.6 Using Molecular Characterization to Make Informed Decisions on the Conservation of Crop Genetic Resources

Information about the genetic make-up of gene bank accessions contributes towards decision-making for conservation activities, which range from collecting and managing genetic resources to identifying genes to add them value for breeding purposes.

Well-informed sampling strategies for germplasm material destined for ex situ conservation and designation of priority sites for in situ conservation are both

crucial for successful conservation efforts. In turn, defining strategies depends on knowledge of location, distribution and extent of genetic diversity. Molecular marker-based characterization, by itself or in combination with other data, such as morpho-phenological traits, provides reliable information for assessing, among other factors, the amount of genetic diversity/similarity, the structure of genetic variation in samples and populations, rates of genetic differentiation among populations and the distribution of biodiversity in populations from different locations.

Molecular characterization helps to determine the breeding behaviour of species, adaptive and reproductive success of individuals and the existence of gene flow among individuals, that is, the movement of alleles within and between populations of the same or related species, and its consequences (Papa and Gepts 2003). Molecular data improve or even allow the elucidation of phylogeny, and provide the basic knowledge for understanding taxonomy, domestication and evolution. As a result, information from DNA markers or DNA sequences offers a good basis for better conservation approaches.

Selected molecular technologies render cost-effective and comprehensive genotype profiles and gene haplotypes of accessions generated through DNA fingerprinting and DNA sequencing that may be used to establish the identity of the material under study. Simultaneously, in addition to the presence of redundant materials or duplicated accessions, these technologies can detect contaminants, and in the case of mixtures, contamination with introgressed genes from other accessions or commercial varieties as well. Moreover, molecular marker data sets provide the baseline for monitoring natural changes in the genetic structure of the accessions as well as changes occurring as a result of human intervention (e.g., seed regeneration or sampling for replanting in the field). Whatever the case, analysis of molecular information allows the design of strategies for either purging the consequences of inappropriate procedures or amending them to prevent future inconveniences (De Vicente et al. 2006).

In the next years, the area of crop plant genomics that might show the greatest development with respect to the use of molecular marker technology is likely that of applied breeding programs.

An increasingly important role of genetic characterization is identifying useful genes in germplasm, that is, maximizing conservation efforts. Because the major justification for the existence of germplasm collections is for the use of the conserved accessions, it is important to identify the valuable genes that can help to breed new varieties able to meet the challenges of current and future agriculture (De Vicente et al. 2006). Characterization has benefited from several approaches resulting from advances in molecular genetics such as Mendelian gene tagging and QTL mapping. Research in this field has led to the acknowledgement of the value of wild relatives, in which modern techniques have discovered useful variation that could contribute to varietal improvement. Knowledge of molecular information in major crops and species, and of the synteny of genomes and colinearity of genes, has also opened up perspectives for identifying important agronomic genes or functional variants in other crop types, particularly those receiving little attention from formal research.

Managing biodiversity means not only genetic characterization through DNA polymorphism detection, as it requires information used to address key issues of both ex situ and in situ plant germplasm management and to assist in the process of decision making (Lanteri and Barcaccia 2006). For ex situ crop germplasm maintenance, molecular tools may contribute to the sampling, management and development of core collections as well as the utilization of genetic diversity. For the in situ and on farm preservation strategies of genetic resources, molecular markers might help in the recognition of the most representative populations within the gene pool of a landrace and the identification of the most suitable strategies for their managing and use.

Analysis of genomic DNA samples based on the detection of molecular markers should be capable of identifying plant varieties unambiguously and definitive and also be effective for calculating the genetic distance between plant germplasm accessions. The ability to identify new varieties and determine their diversity with respect to previously registered varieties promises to be the prime requirement for a valuable market of plant varieties, as well as a guaranteed genetic value of plant materials.

New molecular approaches have been recently developed for adapting the current PCR-based techniques to target functional diversity. New findings from genomics research indicate that there is a tremendous genetic potential locked up in germplasm collections that can be released only by shifting the paradigm from searching for phenotypes to searching for superior genes with the aid of molecular linkage maps (Lanteri and Barcaccia 2006). At present the increasing information available from genome scanning and gene mapping means that molecular markers known to be tightly linked to traits of agronomic interest can be better addressed for characterizing genetic diversity and help in identifying genetic variation of use to breeders. Furthermore, the identification of Mendelian genes and QTLs controlling a given qualitative or quantitative trait and the availability of their DNA sequences may facilitate the classification of variation in germplasm pools. High resolution genetic maps and linkage groups enable markers closely linked to agronomically important traits to be used and the increasing numbers of SNPs and ESTs for genes and transcripts, respectively; provide routes for more targeted sequence-based approaches. Classification of the sequence variants at a target locus would substantially reduce the amount of work needed to assess their potential for breeding and lead to the identification of superior haplotypes, alleles and/or genotypes. The information acquired is now being exploited to transfer different traits, including biotic stress resistances and improved quality traits, to important varieties by means of MAS programs.

Germplasm in collections can undergo structural molecular characterization (i.e., based on the investigation of anonymous DNA sequences) and functional molecular characterization (i.e., based on the identification of genes and their functions). The information gathered from structural characterization not only provides increased clarity on existing genetic diversity and its organization in individuals, but it also useful to determine the structure of populations providing the basis for functional characterization.

The increasing number of genome and transcriptome sequencing projects has opened the opportunity to design functional molecular markers on expressed sequences of known chromosome position for characterizing and exploring genetic resources. Moreover, this information enables the compilation of large amounts of sequence data that can be used to develop markers linked to specific genes and to discover novel functional variations. In addition, the development of technologies continues and this means, on one side, increased markers and, on the other, decreased costs, so to allow their application in the tasks of characterizing and preserving plant genetic resources, which usually involve large numbers of samples.

New developments are also taking place in designing better approaches to access new and useful genetic variation in collections, such as allele mining and association mapping studies. Allele mining focuses on the detection of allelic variation in important genes and/or traits within a germplasm collection. If the targeted DNA is available, either a gene of known function or a given sequence of unknown origin, then the allelic variation in a collection, usually due to point mutations, can be successfully identified. Association mapping studies of natural populations are an alternative to segregation analysis in experimental populations for identifying useful genes by correlation of molecular markers to a specific phenotype. These studies can be performed on a germplasm collection and also on other breeding materials as long as significant linkage disequilibrium (LD) exists.

The importance of the variation captured in genetic resources in allowing evolution and/or facilitating plant breeding has been long recognized. However, appreciating the variation held in collections is not sufficient. Conservation of genetic resources needs to be combined with an enhanced use of conserved materials. Worldwide germplasm collections of crop plant species maintained ex situ in gene banks together with that held in situ and on farm situations harbour abundant quantities of hidden allelic variants. The challenge is to unravel the mysteries of this variation so that it can be used for the benefit of humankind. More and more, technologies have increased throughputs, which generally means the generation of progressively larger amounts of genetic data. Genotyping individuals to identify the available allelic variation that makes up the phenotypes provide the groundwork on which genetic resources can be used in plant breeding. Phenotyping is very much linked to the usefulness of good molecular characterization, together forming the basis of progress in modern genomics research in crop plants (De Vicente et al. 2006).

10.7 Conclusion

In conclusion, the most important challenges in the near future are certainly the molecular characterization of germplasm collections for preserving them from genetic erosion and the identification of phenotypic variants potentially useful for breeding new varieties. Knowing the presence of useful traits, genes and alleles would help in making decisions on the multiplication of accessions and the maintenance of seed stocks for responding to an expected higher demand of materials.

Use of molecular markers for characterization and conservation of genetic resources should be implemented so that genotypes with known and useful genes and alleles can be added to core collections to make them exploitable by breeders. This will facilitate the use of, and add value to, crop plant germplasm resources. A new concept that might be successful is that of building collections primarily based on the knowledge of the presence of valuable genes and traits.

References

Barcaccia G, Albertini E, Torricelli R et al. (1998) Molecular characterization of lentil and emmer landraces. National Conference of Biodiversity, Alghero, Sassari, Italy, 8–11 September, pp 841–846

Barcaccia G, Lucchin M, Parrini P (2003) Characterization of a flint maize (*Zea mays* var. *indurata*) Italian landrace: II. Genetic diversity and relatedness assessed by PCR-based markers. Genet Resour Crop Evol 50:253–271

Barcaccia G, Albertini E, Pallottini L et al. (2005) Fingerprinting wheat varieties for multiple breeding purposes. Proceedings of the XLIX national conference of agriculture genetics, Potenza, 12–15 September, C.06

Barcaccia G, Pallottini L, Parrini P et al. (2006) A genetic linkage map of a flint maize (*Zea mays* var. *indurata* L.) Italian landrace using a one-way pseudo-testcross strategy and multilocus PCR-based markers. Maydica 51:469–480

De Vicente MC, Guzmán FA, Engels J et al. (2006) Genetic characterization and its use in decision-making forthe conservation of crop germplasm. In: Ruane J Sonnino A (eds) The role of biotechnology in exploring and protecting agricultural genetic resources, FAO, Rome. pp 129–138

Dice LR (1945) Measures of the amount of ecological association between species. Ecology, 26:297–302. Proceedings of the international workshop. The role of biotechnology, Villa Gualino, Turin, Italy, 5–7 March 2005. pp 129–138

Ferreira ME (2006) Molecular analysis of gene banks for sustainable conservation and increased use of crop genetic resources. In: Ruane J, Sonnino A (eds) The role of biotechnology in exploring and protecting agricultural genetic resources, FAO, Rome. pp 121–127

Lanteri S, Barcaccia G (2006) Molecular markers based analysis for crop germplasm preservation. In: Ruane J, Sonnino A (eds) The role of biotechnology in exploring and protecting agricultural genetic resources, FAO, Rome. pp 55–66

Lucchin M, Barcaccia G, Parrini P (2003) Characterization of a flint maize (Zea mays var. indurata) Italian landrace: I. Morpho-phenological and agronomic traits. Genetic Res Crop Evol, 50: 315–327

Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J 3:175–182

Negri V, Tosti N, Falcinelli M et al. (2000) Characterization of thirteen cowpea landraces from Umbria (Italy). Strategy for their conservation and promotion. Genet Resour Crop Evol 47:141–146

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci U S A 70:3321–3323

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics, 89:583–590

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci U S A 76:5269–5273

Pallottini L, Garcia E, Kami J et al. (2004) The genetic identity of a patented yellow bean. Crop Sci 44:968–977

Pallottini L (2002) Variation of the genetic structure of populations in relation to different strategies of germplasm conservation (*on farm*, *in situ* and *ex situ*) in local varieties of maize: the case study of the "Nostrano di Storo" (*Zea mays* var. *indurate* L.).Ph.D. thesis, Padua, Italy, University of Padua, pp 316

Papa R, Gepts P (2003) Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. Theor Appl Genet 106:239–250

Rafalski A, Morgante M (2004) Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. Trends Genet 20:103–111

Vendramin G, Morgante M (2006) Genetic diversity in forest tree populations and conservation: analysis of neutral and adaptive variation. In: Ruane J, Sonnino A (eds) The role of biotechnology in exploring and protecting agricultural genetic resources, FAO, Rome, pp 145–147

Vos P, Hogers R, Bleeker M et al (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19:395–420

Wright S (1978) Evolution and the genetics of populations. Variability within and among natural populations. University of Chicago Press, USA, p 4

Part III
Genomics

Chapter 11
Rice Genomics

Gateway to Future Cereal Improvement

Narayana M. Upadhyaya and Elizabeth S. Dennis

Abstract There is an urgent need to increase world's food production to meet the ever increasing demand. Rice plays a direct role as the cereal feeding half the world's population and as an experimental workhorse. It has the smallest genome among cereals, has remarkable similarities with other cereals in sequence, structure, order and function of genes. Foreign genes can be incorporated into rice with ease by genetic transformation. Rice genetics is also well studied and understood. All these features have made rice a model cereal for functional genomics. With the availability of the near complete genome sequence, the emphasis is not only on understanding the functions of the predicted 35,000–50,000 genes, but also unraveling how these genes interact to control important agronomic traits under different environmental conditions. This chapter covers the genome-wide molecular techniques currently being employed in rice with the ultimate goal of achieving much needed increases in crop productivity.

11.1 Introduction

There is a real urgency in enhancing food production to meet the increasing global demand driven by the increase in human population, the reduction in land available for sustainable crop production and the poor performance of crop cultivars under increasingly adverse environmental conditions. Rice, being the major staple food for almost half of the world's population, has naturally been the model cereal for genetic, breeding and agronomic research. More recently, the scientific community also adopted rice as the model cereal for genome research, because of its small genome size (~430 Mb), the ease with which it can be transformed and the similarities of its gene order and gene sequence with other cereals. It is likely that genes identified in rice as being important agronomically will also be important in other cereals thus pave the way for the application of the knowledge gained in rice to other cereals.

N.M. Upadhyaya (✉) and E.S. Dennis
CSIRO Plant Industry, Canberra, Australia
e-mail: Narayana.Upadhyaya@csiro.au

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_11,

Conventional breeding over the last three decades has resulted in a doubling of rice production (Khush 1997). However, breeders are in need of new tools and resources with which they can address the major production constraints such as pests, pathogens, salinity and drought in order to provide the required increase in the rate of production. Rice genomics has the potential to provide such tools and resources in the form of molecular markers for genes and gene control sequences determining the desired traits or as genes and gene control sequences *per se* for use in transformation breeding.

With the near completion of the sequence of both the *japonica* (cv. Nipponbare) and *indica* (cv. 93-11) sub-species, the emphasis now is on determining the function of each of the predicted 35,000–50,000 rice genes. Genome-wide research tools such as data mining for structural similarities, expression profiling at the RNA level with expressed sequence tags (ESTs), oligonucleotide or cDNA chips, expression profiling at the protein level (proteomics), gene knockouts or loss of function studies with naturally occurring alleles, induced deletion and insertional mutants, and gene expression knock-down (gene silencing) studies using RNAi have become integral parts of plant functional genomics. This chapter covers some of the cutting edge molecular techniques currently being employed in rice functional genomics.

11.2 Genome Sequence and Annotation

A map-based, clone-by-clone sequence of the *japonica* cultivar Nipponbare produced by the International Rice Genome Sequencing Project (IRGSP, International Rice Genome Sequencing Project 2005) is now considered to be the “reference” sequence for rice by the international research community. This reference sequence has ~95% coverage of the genome and ~99% coverage of the euchromatic region (see Matsumoto et al. 2007). The Rice Annotation Projects of both the IRGSP and The Institute for Genomics Research (TIGR) are independently assembling and annotating this sequence. Whole genome shotgun sequences of *japonica* cv. Nipponbare and *indica* cv. 93–11 produced by Syngenta (Goff et al. 2002) and the Beijing Genomics Institute (Yu et al. 2002, 2005), respectively, are the other two major rice sequencing efforts. Salient features of these sequences are presented in Table 11.1. The availability of both *japonica* and *indica* sequences is helping to not only improve the accuracy of the rice genome sequence by targeted re-sequencing and gap filling, but also to improve gene annotations and identify biologically significant sequence variations within the *Oryza sativa* sub-species. For example, Beijing Genomics Institute (BGI) researchers have identified an average of 3 single nucleotide polymorphisms (SNPs) in genic regions and 15.1 SNPs in non-genic regions per kilobase between the Nipponbare and 93-11 sequences (Yu et al. 2005). Although some of these could be sequencing artifacts, others could be real allelic variations of biological significance.

The *Oryza* Map Alignment Project (OMAP) initiated by the Arizona Genome Institute (Ammiraju et al. 2006; Wing et al. 2007) aims to create a genome level experimental system for the genus *Oryza* that can be used as a research platform to

Table 11.1 Rice genome sequence assembly and annotation – current status

Assembly	TIGR	IRGSP[b] (RAP-DB)	Syngenta[c]	BGI[d]
Cultivar	Nipponbare	Nipponbare	Nipponbare	93–11
Coverage	10×	10×	6×	6.3×
Genome version	Release 5	Build 4	Release 2	Release 2
Sequenced genome size (bp)	372,077,801	372,089,805	337,414,820[e] (353,378,970)[f]	360,157,649[e] (374,545,499)[f]
Total of 12 chromosomes				
Estimated complete genome size	388,820,000	388,820,000	NA	NA
Unassigned sequences (bp)	NA	NA	76,576,826	104,840,190
Predicted genes including transposable element (TE) genes	56,278 (66,710)[g]	ND	54,296	59,660
Predicted non-TE genes	41,046 (51,286)[g]	53,461	45,824	49,088
Mapped cDNA among 35,187 full-length cDNA	32,775	32,745	31,928	30,354

[a] TIGR = The Institute for Genomic Research. TIGR is now merged with the J. Craig Venter Institute (http://www.jcvi.org/) and TIGR's rice annotation project is moved to Michigan State University (http://rice.plantbiology.msu.edu/).
[b] IRGSP (rap-db) = International Rice Genome Sequencing Project and its Rice Annotation Project database (http://rapdb.dna.affrc.go.jp/).
[c] shotgun sequence from syngenta analysed by Beijing Genomic Institute (BGI).
[d] BGI (http://rice.genomics.org.cn/index2.jsp).
[e] Genome sizes based on the sum total of assigned contigs.
[f] Figures in the parenthesis are the genome sizes as the sum total of genome assigned scaffolds.
[g] Figures in the parenthesis are total genes including splice variants.

study evolution, development, genome organization, ploidy, domestication, gene regulatory networks and crop improvement. To achieve this, the OMAP group is constructing deep-coverage large insert bacterial artificial chromosome (BAC) libraries from 11 wild *Oryza* species and 1 cultivated African *Oryza* species (*O. glaberrima*), fingerprinting and end sequencing the clones, constructing the physical maps and aligning them with the IRGSP reference genome sequence.

The most straightforward way of predicting the function of a rice gene sequence is by comparison with sequence databases from other organisms, as functionally-similar genes normally have sequence similarities at both the protein and DNA level. Supercomputers and robust bioinformatics capabilities are being developed to increase the precision with which sequences can be compared. Several laboratories have

embarked on rice sequence annotation using this approach (see Antonio et al. 2007). Such computational gene predictions suggest that there could be more than 50,000 rice genes with ~60% having some evidence of expression. For example, according to IRGSP's Rice Annotation Project database (RAP-DB; http://rapdb.dna.affrc.go.jp/), among the 53,461 predicted rice genes 31,439 have evidence of expression and 25,012 are protein coding loci with full length cDNA support, others being *abinitio* predictions without any transcriptional evidence. The five categories of open reading frames in IRGSP's RAP-DB are (1) those with 98% identity with known rice proteins; (2) those with 50% identity to known rice proteins; (3) InterPro (http://www.ebi.ac.uk/interpro/) domain-containing proteins; (4) conserved hypothetical proteins; and (5) hypothetical proteins. The current gene numbers belonging to each of these categories are 664, 7,562, 13,632, 6,954 and 1,380 respectively.

Working models for more than 50,000 rice genes have been built by combining available EST data from rice and other plant species. However, among 37,187 full-length cDNAs derived from rice, ~2,400 are yet to be mapped to the rice genome (Kikuchi et al. 2003; Satoh et al. 2007), and the remaining computationally-predicted genes are yet to be confirmed experimentally. Even with so-called full-length cDNAs we have observed occasional unspliced or incorrectly spliced intron sequences. There are also some discrepancies in the number of mapped full- length cDNAs, mapping strand and mapping chromosome between the TIGR, IRGSP, BGI Syngenta and BGI 93-11 assemblies. Such differences between *japonica* Nipponbare and *indica* 93-11 genomes are expected because of accumulated independent gene duplications and inter- and intra-chromosomal rearrangements over the millions of years of divergence that have occurred between these two *Oryza* sub-species.

It is perhaps worthwhile noting that although the Nipponbare genomic sequences used in TIGR and IRGSP/RAP-DB curations are the same (map-based sequence), the way they are assembled chromosome-wise into 12 pseudomolecules, further annotated and curated are different. In RAP-DB an ID (OsXXg#######) consists of the species name (Os for *Oryza sativa*), a two-digit number for chromosomes, the type identifier (g for genes), and a seven-digit number that indicates a sequential order of loci along a given chromosome. The RAP-DB annotations with the Os identifiers are in the DDBJ/EMBL/GenBank under the accession numbers (for the 12 latest build 4 pseudomolecules) AP008207 to AP008218 (http://www.ddbj.nig.ac.jp/ddbjnew/mag/mag28-e.html#02) and corresponding NCBI (National Centre for Biotechnology Information) Reference Sequence (http://www.ncbi.nlm.nih.gov/RefSeq/) accession numbers are NC_008394 to NC_008405. TIGR IDs are in the form LOC_OsXXg##### and were determined by a different method based on TIGR's own genome assembly. Thus, some of the RAP-DB transcripts corresponding to TIGR transcripts may not have the same nucleotide sequence. Other differences are (1) physical gaps are represented by 1,000 Ns in TIGR and 100 Ns in RAP-DB besides denoting ambiguous sequences as "N"; (2) assigned consecutive locus IDs are in multiples of 10 (e.g., LOC_Os01g01010, LOC_Os01g01020…) in TIGR and are in multiples of 100 (e.g., Os01g0100100, Os01g0100200…) in RAP-DB; and (3) TIGR has the provision for designating alternative spice forms (e.g., LOC_Os01g01010.1, LOC_Os01g01010.2…).

Efforts to obtain the remaining 5% of the rice genome sequence, comprising mostly telomeres and centromeres, are currently underway (Ammiraju et al. 2005; Mizuno et al. 2006a, b). Nevertheless, with the available *japonica* and *indica* genome sequences, attempts to unravel allelic variations between these two sub-species can be made with high confidence using various approaches, some of which are highlighted in the following subsections.

11.3 RNA Expression Studies

The first product of a gene is single stranded RNA, transcription of which is thought to be controlled by complex and highly coordinated processes. RNAs are also subjected to post-transcriptional control in each cell or tissue throughout the life cycle of the organism. Although there could be more than 50,000 rice genes, not all of these will be transcribed at any given time, tissue or environmental condition and in addition some of the RNA will be suppressed, broken down or non-translatable. Thus, one challenge is to characterize all of the transcribed regions of the genome. Such transcriptome analysis is normally attempted by collecting large numbers of ESTs from diverse cDNA libraries. To date, there are 1,220,876 rice ESTs in the public database (http://www.ncbi.nlm.nih.gov/dbEST/). Recent advances in the technology for construction of full-length cDNA libraries have made it possible to produce more than 30,000 rice full-length cDNAs (Kikuchi et al. 2003; Satoh et al. 2007). This has helped in improving rice genome annotation, gene organization and genome-wide expression profiling. One other significant EST and full-length cDNA collection from *indica* rice comes from BGI which can be viewed through BGI-RIS (Zhao et al. 2004; http://rice.genomics.org.cn/index2.jsp).

Genome-wide expression profiling (including differential expression) of rice genes is being facilitated by high throughput techniques such as microarrays, Serial Analyses of Gene Expression (SAGE), Massively Parallel Signature Sequencing (MPSS) and more recently by ultra deep sequencing (e.g., 454, Solexa and SOliD). Spatial and temporal RNA expression patterns could provide an insight into their cellular and developmental functions. Most of the microarray data produced by different groups and the experimental conditions used to produce these data can be accessed from public databases such as NCBI-GEO (http://www.ncbi.nlm.nih.gov/projects/geo/), Array Express (http://www.ebi.ac.uk/microarray-as/ae/) and Rice Expression Database (http://red.dna.affrc.go.jp/RED/). Following the first published gene expression profiles of abscisic acid- and gibberellin-responsive genes in rice (Yazaki et al. 2004) using the 22K rice oligomicroarray (produced by NIAS in collaboration with Agilent Technologies), several research groups have produced microarray data to study rice gene expression under diverse experimental conditions. Different arrays used include the NSF Rice Oligonucleotide Array V2 20K and V3 45K (http://www.ricearray.org/index.shtml), Agilent 22K (now obsolete) or 44K custom arrays (Agilent Technologies), Affymetrix 51K GeneChip Array (http://www.affymetrix.com/products/arrays/specific/rice.affx) and the 70-mer

microarray 42K array (Ma et al. 2005). Using the Agilent 44K microarray Shimono et al. (2007) have identified benzothiadiazole (BTH)- and salicylic acid (SA)-inducible WRKY transcription factor (TF) genes and confirmed the important role of WRKY45 in benzothiadiazole-inducible blast resistance in rice.

A variant of microarray technology, the genome tiling microarray, has been employed to evaluate computationally annotated rice gene models and to identify hitherto unidentified transcription units (Li et al. 2005, 2006). Genome tiling arrays involve the generation of a virtual tiling path covering a target genomic region. They are made up of short oligonucleotide probes immobilized on the surface of glass slides or Affymetrix chips. Hybridization with fluorescence-labelled RNA samples generates signals that reflect the transcriptional activities of the target genome which are compared with the available genome annotation data. Rice genome tiling microarrays have been developed based on the Markless Array Synthesizer (MAS) technology and contain 36-nt oligonuclotide tiling of both the *japonica* and *indica* genome sequences with an average 10-nt space (Stolc et al. 2005). Using this approach Li et al. (2006) detected the transcription of 35,970 *indica* rice gene models out of the computationally predicted 43,914 non-TE protein coding gene models (Yu et al. 2005). They detected antisense transcription of 19.6% of 3,019 *japonica* chromosome 10 gene models (Li et al. 2005) and 23.8% of 43,914 *indica* gene models (Li et al. 2006). They also detected 5,464 unique novel transcriptionally-active regions (TARs) in the intergenic regions.

Besides quality control and hybridization reproducibility issues, microarray data analysis and interpretation are challenging because of the sheer quantity of data to be handled and the large numbers of genes showing differential expression. Moreover, some genes showing large expression differences could have minimal impact while others with minimal expression differences could have profound effects. Genes showing marginal differential expression are easily overlooked in microarray analyses. Microarray expression data should be validated by other means such as quantitative RT-PCR or Northern blot hybridization.

Some of the inherent limitations of expression profiling by ESTs (such as high cost, high labour and inability to perform tissue specific expression quantification) and microarray (such as background signal intensities rivalling those of weakly expressed mRNAs, inability to distinguish closely related transcripts or transcript variants and incomplete genome coverage) can be overcome by tag-based technologies such as SAGE and MPSS (see Kikuchi et al. 2007). SAGE was the first tag-based method developed for qualitative and quantitative profiling of genes which requires no prior information (Velculescu et al. 1995). In SAGE, essentially short sequence tags (14–15 bp) are produced by special restriction enzymes (e.g., *Nla*III), made into ditags and then concatenated (70–100 ditags), cloned, sequenced and individual tags identified and computed. Thus, the data obtained (tags and their frequency) reflects the actual gene expression pattern at the time of RNA isolation, allowing visualization of transcriptome complexity, including alternative splicing, antisense transcripts and 3’ UTR variants (Patankar et al. 2001; see Kikuchi et al. 2007). One inherent limitation of using short tags is in their assignment to duplicated or repeat sequences (Chen et al. 2000). Several improvements such

as LongSAGE in which a different tagging RE (*Mme*1) is used which produces 21-bp tags (Saha et al. 2002) or SuperSAGE in which the tagging RE (*Eco*P151) produces 26-bp tags (Matsumura et al. 2003) have been made. The latter has been used effectively to identify 5' tags of expressed genes by cap analyses of gene expression (CAGE) by several researchers (see Kikuchi et al. 2007). Recently, Gowda et al. (2004, 2007a) have developed a Robust LongSAGE method and successfully applied this method to analyse the defence associated transcriptome of rice. In their study 13.1% (7,597) of tags matched to the *M. grisea* genome sequence and only 7.1% matched to TIGR *M. grisea* ESTs to suggest the presence of novel transcripts expressed only during pathogen infection. They also found a good number of antisense tags suggesting the presence of non-coding antisense transcripts or small RNAs as discussed in the following section. These data can be accessed on the *Magnaporthe grisea Oryza sativa* (MGOS) database (http://www.mgosdb.org/sage/).

Before the advent of other deep sequencing technologies, MPSS was another powerful tag-based technology. It was developed by Brenner et al. (2000) and commercialized originally by the company Solexa Inc (now owned by Illumina Inc.) and has been applied in deep transcriptome analyses of *Arabidopsis* (Meyers et al. 2004a, b), rice (Nobuta et al. 2007; Lu et al. 2008) and other plant species. Although sequence signatures obtained are similar to those from SAGE, MPSS uses a novel cloning and sequencing method wherein tag sequences are obtained simultaneously by sequencing off the beads using a technique of enzymatic digestion and hybridization (Nobuta et al. 2007). Nobuta et al. (2007) sequenced ~47 million signatures from 22 cDNA libraries. The data showed sense strand expression of 25,500 annotated rice genes and antisense expression of 9,000 annotated genes with another 15,000 signatures mapping to unannotated genomic regions. These unannotated TAR specific transcripts and antisense transcripts most likely are involved in the biogenesis of different classes of small RNAs discussed in the following section.

Ultra deep sequencing technologies such as 454, Solexa and SOliD are set to revolutionize genome sequencing and annotation. The 454 sequencing-by-synthesis (SBS) is based on the detection of nucleotide incorporation in a primer-directed polymerase extension. The sequence can be deduced iteratively based on the pyrophosphate released during the DNA polymerase reaction, the quantitative conversion of pyrophosphate to ATP by sulfurylase, and the subsequent production of visible light by firefly luciferase (Ronaghi et al. 1996). Solexa, Inc. has developed a novel four-colour DNA SBS method using dye-termination chemistry (http://www.solexa.com), with a potential of producing >10 million 25-nt to 30-nt sequence tags with high accuracy. Another sequencing approach, SOLiD (Supported Oligo Ligation Detection), is being developed by Applied Biosystems, Inc. Here an array of microbeads each coated with a single DNA or cDNA fragment and a pool of fluorescent oligos are used to "read" the sequences by complementary binding with a repeated process of ligation, detection and cleavage. This method can produce 50-nt sequence per bead in parallel for >10 million beads.

With all these new developments in deep sequencing technologies we are seeing an explosive increase in the RNA expression tag and small RNA datasets from

diverse plants, environmental conditions and experimental treatments which will help in unravelling the complexities of the transcriptome, including that of non-coding RNAs in diverse biological systems.

11.4 Small RNA Studies

Short non-coding RNA molecules are now emerging as important regulatory molecules involved in both transcriptional and post-transcriptional suppression of gene expression in eukaryotes (see Phillips et al. 2007). Although precise pathways in plants are still to be fully understood, these small RNAs act by inhibiting translation, cleaving mRNA or by chromatin regulation. There is growing evidence to suggest that small RNAs play a major role in plant response to environmental stresses (Borsani et al. 2005; Phillips et al. 2007; Sunkar et al. 2007). Large datasets of small RNAs are being generated by several laboratories using ultra deep sequencing technologies (454, Solexa, SOLiD). Bioinformatics tools required for processing these large datasets and for predicting different classes of small RNAs are being developed (Xu et al. 2008; Zhu et al. 2008).

On the basis of their mode of biogenesis and the structure of the genomic loci from which they are transcribed, small RNAs are classified as microRNAs (miRNAs) and endogenous small interfering RNAs (siRNAs). MicroRNAs are predominantly 21 nt in length and are derived from 70- to 500-nt long single stranded primary transcripts (pri-miRNAs) by the action of RNase III-like enzymes DICER-LIKE1 (DCL1) or DCL4. The mature miRNA is loaded to the RNA induced silencing complex (RISC) to guide the complex to the target mRNAs (containing a stretch of perfect or near perfect complementary sequence). Most plant miRNAs cleave the target mRNA, however evidence is mounting to suggest that some plant miRNAs suppress translation without cleavage. Since the first report of miR395, a novel *Arabidopsis* miRNA predicted to target ATP sulfurylase genes involved in inorganic sulphate assimilation (Jones-Rhoades and Bartel 2004), several highly conserved miRNAs and their targets have been computationally identified (see Phillips et al. 2007). In addition to such conserved miRNAs, novel non-conserved species-specific plant miRNAs have been identified. The latest information on plant miRNAs can be obtained from the miRBase database maintained by Sanger Institute (http://microrna.sanger.ac.uk; Griffiths-Jones et al. 2008).

Endogenous siRNAs are further classified into *trans*-acting siRNA (ta-siRNA), heterochromatic siRNA (hc-siRNA) and natural antisense siRNA (nat-siRNA). These siRNAs are produced from long dsRNA molecules generated by RNA-dependent RNA polymerases (RDRs). In general, hc-siRNAs and nat-siRNAs act in *cis* to target the locus they are derived from whereas ta-siRNAs target mRNAs from different loci. Ta-siRNA gene transcripts are initially targeted by miRNAs and the cleaved transcripts are transcribed into double stranded by RDR6 before further in-phase cleavage into 21-nt ta-siRNAs occurs via the action of DCL4. These ta-siRNAs then guide the appropriate RISCs (AGO1 or AGO7) to target cognate mRNAs.

In nat-siRNA biogenesis, a 24-nt nat-siRNA is produced from the overlapping antisense transcripts of two transcriptional units, by the combined action of the silencing proteins DCL2/RDR6/SGS3/NRPD1a which then targets one of the overlapping mRNAs. This first cleavage triggers the DCL1 pathway to produce several 21-nt siRNAs to target the same mRNA. The hc-siRNAs are mostly 24 nt in length and are produced by DCL3 from primary transcripts generated by the plant specific RNA polymerases (PolIVa and/or PolIVb) after conversion into dsRNA by RDR2. This subclass of siRNAs are thought to be involved in transcriptional silencing of several types of repetitive sequence, including retroelements and 5S rDNA arrays, by DNA methylation and heterochromatin formation (see Phillips et al. 2007).

Among the ~3 million MPSS derived small RNAs from inflorescence, seedling and stem libraries the majority of low abundance small RNAs were mapped to repetitive sequences, intergenic and genic regions (Nobuta et al. 2007). However, several clusters of highly regulated small RNAs were also observed (http:mpss.udel.edu/rice). From three more small RNA libraries (untreated seedlings, seedlings treated with abscisic acid or with the rice blast pathogen *M. grisea*), Lu et al. (2008) obtained ~4.5 million 17-nt signatures. These contained signatures of 24 new miRNA genes and another unique class of miRNAs that derived from natural *cis*-antisense transcript pairs. These natural miRNAs (nat-miRNAs) have large introns in their precursor transcripts and require DCL1 activity for their biogenesis.

To investigate the roles of miRNAs in rice grain development Zhu et al. (2008) carried out deep sequencing (454 and Solexa) of the small RNA population of rice grains at two developmental stages. In a dataset of ~5.5 million sequences, in addition to representatives of all 20 conserved plant miRNA families, 39 novel rice miRNA families expressed specifically in grains were identified. Predicted target mRNAs for a number of these new miRNAs were confirmed by cleavage assays, with the authors presenting the first evidence for miRNA and miRNA-like small RNA expression in an organ-specific manner. In another study, Morin et al. (2008) compared the small RNA transcriptome of *Pinus contorta* with that of *Oryza sativa* by 454 deep sequencing. They found predominance of 21-nt small RNAs in *P. contorta* in contrast to rice where both 21-nt and 24-nt small RNAs were of similar abundance. They found 18 highly conserved known miRNA families as well as numerous small clusters of conserved small RNAs of unknown function. Using several approaches including expressed sequence folding and machine learning algorithms, they found 51 candidate novel miRNA families specific to *P. contorta*. Six perfectly conserved classes of small RNAs between *P. contorta* and rice were also detected suggesting that the RNA silencing processes were highly developed in the earliest spermatophytes.

From small RNA libraries of control, drought stressed and salt stressed rice seedlings, Sunkar et al. (2008) obtained 58,781, 43,003 and 80,990 unique genome-matching small RNAs by pyrosequencing. They detected most of the conserved rice miRNAs as well as 23 novel miRNAs, each originating from a unique locus in the rice genome. Six of the newly identified miRNAs were shown to be conserved in other monocots. An additional 40 candidate miRNAs were identified and for 9 of them they could predict 20 different targets.

Overlapping genes (antisense transcripts overlapping with protein coding transcripts) are ubiquitous in the plant and animal kingdoms with 4 to 20 percent of genes belonging to this category. Such natural antisense transcripts (NATs) have been shown to be involved in alternate C4 metabolism, alternative splicing, RNA editing, DNA methylation, genome imprinting, X-chromosome inactivation and in the regulation of circadian clock genes (see Lapidot and Pilpel 2006). The association of NATs and the involvement of NAT-derived siRNA (nat-siRNA) in the control of salt stress responsive gene expression in *Arabidopsis* (Borsani et al. 2005) and *Magnaporthe grisea* infection stress responsive gene expression in rice (Gowda et al. 2007b) have recently been demonstrated. We have carried out expression studies on 3,174 bioinformatically-predicted rice NAT genes under water stress and non-stress conditions using a custom built (Combimatrix) oligoarray and identified several NAT gene pairs showing anti-correlated expression. However, such expression could be only confirmed for a few NAT gene pairs by qRT-PCR. To investigate the production of NAT-specific and other small RNAs in response to stress we carried out deep sequencing (Solexa) from drought stressed and non-stressed plant samples. Among 4.8 and 4.0 million reads from stressed and non-stressed samples respectively, 1.64 and 1.30 million reads were unique small RNAs with the majority (>67%) being either 21 nt or 24 nt in length. Among the 133,235 rice NAT-specific 18-nt to 24-nt unique small RNAs, 4,405 mapped to the overlapping region of the NAT pairs, of which 63 were stress responsive and 35 of these mapped to retrotransposon related genes. In the dataset we have also identified several known and new miRNAs which appear to be drought stress responsive (Upadhyaya et al. unpublished).

11.5 Proteomics

Information on RNA expression alone will not always shed light on gene function, interaction and regulation. The majority of gene products, including proteins, do not act in isolation. The challenges for rice researchers are identification of the function, regulation, protein-protein interactions, and type of post-translational modification of each encoded protein. Also, the dynamic nature of the proteome response to external and internal cellular events, in the form of relative abundance changes, post-translational modifications and interactions with other proteins, makes it very challenging to study. During the past couple of years considerable progress has been made in the systematic, functional characterization of proteins in the various tissues and organelles of rice (see Komatsu 2007). A system for direct differential display using two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) (O'Farrell 1975) has been developed for the identification of rice proteins that vary in expression under different physiological conditions and among different tissues.

Several public databases of 2D-PAGE derived plant proteins are already available, such as WORLD-2DPAGE (http://expasy.org/ch2d/2d-index.html), Rice

Membrane Protein Library (http://www.cbs.edu/rice/) and the Rice Proteome Database website (http://gene64.dna.affrc.go.jp/RPD/) which provide extensive information on the progress of rice proteome research. In the 2D-PAGE based approach, intact proteins are separated by 2D-PAGE and protein abundance is determined by stain intensity of the protein spot on the gel. The differential proteome is confirmed by image analysis. The identity of the protein is generally determined by MS analysis of peptides after proteolysis of the protein spot or by protein sequencing after blotting the gel to a membrane. The 2D-PAGE based approach has been routinely used for large-scale quantitative proteomics analyses. Differential proteomes are available for various stresses (cold, drought, salinity, ozone, fungal infection, and viral infection) and hormone treatments (gibberellin, brassinosteroid, jasmonic acid, auxin). With fluorescence 2D difference gel electrophoresis (2D-DIGE), one could resolve up to 1,500 proteins (Komatsu et al. 2006).

In addition, progress is being made in detecting post-translational modifications such as glycosylation, lipid attachment, phosphorylation, methylation, disulfide bond formation and proteolytic cleavage. For example, Khan et al. (2005) carried out phosphoproteome analyses of different rice tissues in response to hormone stresses using in vitro radiolabelling (^{32}P) and MS analysis. They showed changes in the phosphorylation status of six proteins in response to gibberellic acid (GA), including glyceraldehyde-3-phosphate dehydrogenase and cytoplasmic malate dehydrogenase, 3 proteins in response to brassinolide (BL) and 5 proteins in response to 2-4-dichlorophenoxyacetic acid (2-4-D). Glyceraldehyde-3-phosphate dehydrogenase and cytoplasmic malate dehydrogenase are involved in the synthesis of various metabolites and the subsequent production of energy. The enhanced phosphorylation of these proteins in response to different hormones indicates that this may be the mechanism through which hormones activate metabolic pathways in rice leaf sheath and thus stimulate plant growth (Khan et al. 2005).

Enzymes often associate each other (transiently or stably) into large protein complexes to increase their efficacy an/or specificity, and in such cases studying protein-protein interactions is equally important (Eubel et al. 2005). Approaches used for such studies include yeast two-hybrid systems, co-immunoprecipitation, pull-down assays, and in vivo fluorescence techniques. However these techniques fail to give a global view of protein-protein interaction. Recently, Lonhosky et al. (2004) have developed hierarchical and non-hierarchical statistical methods to analyse the expression patterns of 526 protein spots on 2-D gels using de-etiolated maize chloroplasts as a model system. Quantitative comparisons with cluster analysis of expression patterns over multiple samples could shed light on to the functions of proteins with no known function. For example, Tanaka et al. (2005) measured changes in the rice seedling's basal region proteins at five time points after sowing using the Mathematical Gene Interaction Network Optimization software (Minos). The software employs the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) clustering utilizing the S-system differential equation formula (Tanaka et al. 2005) and estimates the cluster interaction by a set of differential equation coefficients that simulate the time course. In this study, a previously unknown

protein was up-regulated along with Cluster 13 which also contained antifungal protein 2, and with Cluster 15 containing fructokinase. The same protein was down-regulated along with Cluster 30 which also contained RuBisCO binding protein α subunit and 60S ribosomal protein L19. This protein expression was regulated oppositely to Cluster 29 containing ADP-ribose pyrophosphatase. These results suggest that this unknown protein is related to pathogen defence, sugar metabolism, and protein maintenance.

Thus the differential display of proteins with 2D-PAGE is a powerful approach to study the complex patterns of protein expression over the course of development. Cluster interaction analyses of these expression patterns, based on the S-system, will be very useful in the identification of protein functions. This approach will be applied to resolving the interactions between proteins and might lead to identifying the roles for proteins involved in rice plant development (see Komatsu 2007).

11.6 Metabolomics

Metabolomics is the comprehensive analysis of low molecular weight compounds in biological samples and is emerging as a biochemical phenotyping tool along with transcriptomics and proteomics in functional genomics (see Tarpley and Roessner 2007). The diverse chemical nature of the metabolome warrants a range of methodologies for extraction, separation, detection and quantification. Most of the commonly used technologies are based on chromatographic separation of complex compound mixtures, using either liquid or gas chromatography and mass spectrometric detection. Nuclear magnetic resonance (NMR) spectroscopy is also playing a major role in metabolomics approaches. Fourier-transform ion cyclotron MS (FT-ICR-MS) can mass-resolve metabolites with a mass accuracy <1 ppm (see Dunn et al. 2005), and thus provides a high throughput method for metabolite fingerprinting. This technique has been used in conjunction with proteomics for comparing leaves, panicles and calli of wild-type rice (cv. Nipponbare) and a transgenic line over-expressing the YK1 gene (homolog of maize HC-toxin reductase gene). The YK1 gene is known to confer increased tolerance to rice blast and multiple environmental stresses (Takahashi et al. 2005). Although global composition of organ-specific metabolites did not differ significantly between the two lines, alterations in less than 10% of the metabolites were observed. The transgenic line expressed several previously reported stress-responsive proteins suggesting that ectopic over-expression of a single gene (YK1) can affect expression of unrelated proteins and metabolites.

Metabolites in different tissue sections obtained from various positions along the developing rice seedling (cv. IR36) and at a number of post-emergence days were determined using GC-MS by Tarpley et al. (2005). They identified a small subset of metabolite biomarkers specific to the period of rice development bridging first tillering. Because this is an important and biologically representative developmental stage, the set of “biomarker” metabolites found could be used for comparative

study of the pattern of metabolite change due to development, environment or genotype. Tarpley et al. (2005) validated these biomarkers by comparison with the diurnal data obtained by Sato et al. (2004).

he current limitations of routine metabolomic analyses as a part of rice functional genomics programs include the access to expertise in the techniques of plant metabolomics, the necessary instrumentation, the establishment of adequate sample preparation procedures, the need for coordinated comprehensive cataloguing and/or control of environmental variables, the availability of databases providing storage and access to metabolome-specific data (see Tarpley and Roessner 2007).

11.7 Natural and Induced Variants

Breeders have been exploiting natural variation for centuries to harness valuable traits such as disease and pest resistance and cytoplasmic male sterility (Brar and Khush 1997). Wild ancestors have been shown to contain valuable alleles contributing to the enhancement of complex traits such as yield (Tanksley and McCouch 1997). Chromosomal regions that are associated with such quantitatively inherited phenotypes (quantitative trait loci or QTL) contain many genes interacting with each other (G × G or epistasis) and with the environment (G × E) to exhibit the trait (see Iyer-Pascuzzi et al. 2007). Isolation of these genes could be achieved using positional cloning strategies (Yano 2001) and such an approach requires dense genetic maps with many visible and molecular markers.

Following the development of the first rice Restriction Fragment Length Polymorphism (RFLP) map (McCouch and Kochert 1988), map density has been increased using both RFLP and simple sequence repeat (SSR) markers (see Iyer-Pascuzzi et al. 2007). For ease of use, 3,267 RFLP markers have been converted to Cleaved Amplified Polymorphic Sequence (CAPS) markers (Yano et al., http://rgp.dna.affrc.go.jp/publicdata/geneticmap2000/index.html) which allow PCR based detection instead of the more cumbersome technique of Southern blot hybridization. The physical maps with overlapping BAC or yeast artificial chromosome (YAC) large insert clones and the subsequent near completion of rice genomic sequencing has greatly accelerated positional cloning. However, the production of mapping populations and the fine mapping of mutant loci are still time consuming. Mapping populations of doubled haploid, F_2, advanced backcross, nearly isogenic, recombinant inbred, chromosome segment substitution, introgression and QTL-NIL have been produced by different research groups (see Iyer-Pascuzzi et al. 2007), and are being utilized for genetic mapping. For example, using mapping populations derived from two rice cultivars (cvs. Kasalath and Nipponbare) differing in photoperiod sensitivity and flowering time, Yammamoto et al. (2000) demonstrated epistasis between *heading date6 (Hd6)* and *Hd2*. The green revolution gene *semi-dwarf1* (*sd1*) was identified using both a natural variant (deletion mutant) and an induced mutant (point mutation) both being knockout mutations in the gibberellin biosynthetic gene *OsGA20ox2* (see Iyer-Pascuzzi et al. 2007).

Genetic variability required for plant breeding research can also be generated by chemical and ionizing radiation mutagenesis. In fact, more than 500 rice lines have been generated using induced mutagenesis and released as new varieties (see http://www-mvd.iaea.org/MVD). Because of the high density of mutations, genome-wide saturation mutagenesis can be achieved using a relatively small mutant population (Koornneef et al. 1982; see Bhat et al. 2007). Such a population has the potential to provide a large allelic series as a complement to the knockout mutants produced by insertional mutagenesis or gene silencing techniques. Recent rapid developments in high throughput screening techniques to identify point mutations such as “Targeting Induced Local Lesions In Genomes (TILLING)” have given further impetus for generating such mutant stocks as a resource for rice functional genomics (see Bhat et al. 2007). Commonly used chemical mutagens include ethylmethane sulfonate (EMS), diepoxybutane (DEB), N-Methyl-N-nitrosourea (MNU), sodium azide, and commonly used irradiation mutagens include fast neutron, gamma and X-rays, accelerated ions.

The current worldwide collection of mutant stocks include a fast neutron-induced mutant stock (24,660 M_2 lines) of *japonica* variety M202 in the United States (Li et al. 2001), fast neutron, gamma ray, DEB and EMS induced mutants of *indica* cv. IR64 (~45,000 M_4 lines) at the International Rice Research Institute (Leung et al. 2001; http://www.iris.irri.org), EMS induced rice lines (60,000 lines) at the Institute of Plant Physiology and Ecology of Shanghai Institute for Biological Sciences (http://202.127.18.254/research/field3.html), sodium azide-induced mutant stock (2000 M2 lines) at Taiwan Agricultural Research Institute (http://www.agnet.org/library/article/rh2003009b.html) and MNU induced mutant stock of cvs. Kinmaze and Taichung 65 (6,000 lines) (http://www.shigen.nig.ac.jp/rice/oryzabase/nbrp-Strains/kyushuGrc.jsp). Both forward genetic (phenotyping, map-based cloning, genome-wide chip based single-feature polymorphism detection) and reverse genetic (PCR based screening and TILLING) approaches are being employed in rice gene identification. The Seattle TILLING Project (http://tilling.fhcrc.org:9366/) in collaboration with IRRI and the Agricultural Research Station of the US Department of Agriculture at Davis is actively developing the TILLING method for rice (Till et al. 2007). The technical challenge for rice TILLING is in achieving the required mutation density. Till et al. (2007) appear to have achieved a frequency of 1 mutation per 300 kb in mutant populations generated using EMS and a combination of sodium azide and MNU. Other technical improvements in rice TILLING include the use of non-labelled primers (instead of fluorescence labelled primers), in conjunction with capillary gel electrophoresis (Suzuki et al. 2005) and a simple agarose gel method which obviates the need for labelled primers and/or special genotyping platforms (Raghavan et al. 2007). IRRI researchers are currently employing TILLING for specific traits such as disease resistance or drought tolerance and are using phenotype enriched mutant subsets for TILLING.

Once a gene responsible for the phenotype under study is identified it is possible to develop “perfect markers” to distinguish favourable from unfavourable alleles. Perfect markers allow breeders to easily and reliably predict the plants with desired phenotype in advance thus reducing the time and cost associated with elaborate

screening. Examples of the successful perfect marker applications are in screening for blast and bacterial blight, red pericarp, grain amylose content and aroma (see Iyer-Pascuzzi et al. 2007).

11.8 Insertional Mutants

Insertional mutagenesis provides a rapid and direct way to clone mutated genes. As the sequence of the inserted element is known, the gene in which it is inserted can be easily recovered using various cloning and PCR-based strategies. Most commonly used random insertion mutagens are retrotransposons (*Tos17*), T-DNA, and transposons *iAc*/*Ds* and *En*/*I* (see Hirochika et al. 2004). With a population saturated with insertions, i.e., having at least one insertion in each gene, it is possible to apply both "forward genetic" and "reverse genetic" approaches to identify gene function. In the forward genetics approach, a mutant with a phenotype is first identified by screening this population, and sequences flanking the insert are then cloned and compared with database sequences to enable assignment of function to the mutated gene. In the reverse genetics approach, one starts with a computer predicted gene from the genome sequence and searches for an insertion mutant in that gene. Oligonucleotide primers from the insertional element and from the gene of interest are used for PCR amplification. Appropriately pooled DNA samples are used for high throughput screening for this often rare event in such populations. Once a mutation in the appropriate gene has been identified homozygotes are isolated and the phenotype confirmed.

Four types of active endogenous transposable elements identified to date in rice are (1) long terminal repeat (LTR) retrotransposons such as *Tos10*, *Tos17* and *Tos19* (Hirochika et al. 1996); (2) long interspersed nuclear element (LINE)-type retrotransposon such as *Karma* (Komatsu et al. 2003b); (3) miniature inverted repeat transposable elements (MITEs) such as *miniature Ping* (*mPing*) (Jiang et al. 2003), and; (4) the *nonautonomous DNA-based active rice transposon* (*nDart*) (Fujino et al. 2005). All of these native rice transposable elements are dormant under normal conditions and become active during tissue culture or after treatment with inducing agents such as γ-irradiation. Low copy number, controllable amplification through the length of tissue culture, unlinked transpositions and preferential insertion into low copy number genes makes *Tos17* an ideal tool for saturation mutagenesis (Hirochika 2001). Nearly 50,000 *Tos17* insertion lines with each line containing ~10 copies of *Tos17* have been generated (Hirochika et al. 2004). However, only 3% of visible mutants could be traced back to a *Tos17* insertion suggesting difficulties in using this method for forward genetics. To date, more than 25 genes have been identified using *Tos17* mutants by a forward or reverse genetics approach (see Zhu et al. 2007). More than 15,000 *Tos17* Flanking Sequence Tags (FSTs) have been categorized for search of insertions in genes of interest (http://tos.nias.affrc.go.jp/).

With the vast improvements in rice tissue culture methodology and transformation efficiencies, T-DNA has emerged as the preferred insertion mutagen for generating

large libraries of insertional mutant lines. Several research groups in Korea, China, France and Taiwan (An et al. 2003, 2005; Chen et al. 2003; Sallaud et al. 2004; see Guiderdoni et al. 2007) are generating T-DNA insertion libraries, characterizing T-DNA flanking sequences at insertion points and gathering phenotypic information in web accessible databases (see Guiderdoni et al. 2007). Todate, more than 460,000 T-DNA lines and ~118,000 FSTs have been produced. Although these resources will be useful for reverse genetics, there are limitations in obtaining T-DNA insertions into smaller genes such as single exon genes which may account for up to 40% of the genes in rice.

Since the first report showing the activity of the autonomous *Ac* element in transgenic rice 18 years ago, sophisticated transposon tagging systems have been developed to improve both tagging and screening efficiencies in rice (see Zhu et al. 2007), and are primarily based on the two-component *iAc*/*Ds* (Chin et al. 1999) or *En/I* (Greco et al. 2004) systems. Since the successful cloning of a gene (*BFL1*/*FZP*) which mediates the transition from spikelet to floret meristem (Komatsu et al. 2003a; Zhu et al. 2003), several genes have been cloned by transposon tagging. Insertional mutagenesis by transposons has distinct advantages over that by T-DNA. Large-scale transposon mutagenized populations can be produced using a relatively small number of starter lines, as many independent insertions can be generated among the progeny of a single parental line. The tagged gene can be confirmed by revertants resulting from excision of the transposon. Transposons can also be remobilised to produce new insertion lines in order to target genes in a specific chromosomal region, i.e., corresponding to mapped QTL. The *iAc*/*Ds*-based gene and enhancer trap systems in rice yield 5–10% unique stable insertion lines (see Zhu et al. 2007). The *Ds* re-insertions linked to the original location of *Ds* within the T-DNA (the *Ds* launch pad) varied from 36–67% with the majority being within one cM of the *Ds* launch pad (seel Zhu et al. 2007).

Most of the T-DNA and transposon (*Ds* or *I*) constructs used as insertional mutagens have been modified to act as gene traps or enhancer traps. Use of a reporter gene with a minimal promoter (enhancer trap) or with intron splice acceptors (gene trap) linked to T-DNA or *Ds* transposon sequences facilitates "trapping" of genetic regions which do not have a visible phenotype and will report on the expression of the chromosomal gene at the site of insertion. For example, expression of the reporter gene in roots will identify sequences which direct root expression. Gene trapping efficiencies of ~6% have been reported for these constructs in rice (see Hirochika et al. 2004). The efficiency of T-DNA gene trapping depends on the frequency of "clean" T-DNA insertions, i.e., insertions devoid of direct or inverted T-DNA repeats or of the incorporation of vector backbone (VB) sequences derived from outside the T-DNA borders (Sallaud et al. 2004; Upadhyaya et al. unpublished). A "clean" *Ds* containing T-DNA is also essential for satisfactory mobilization of *Ds*. We have developed a novel method of producing stable *Ds* insertion lines using a transiently-expressed transposase (TET) system (Upadhyaya et al. 2006). We have developed constructs suited for high efficiency insertional mutagenesis in general, and the TET system in particular. By super-infecting callus tissue from single-copy *Ds*/T-DNA lines, having both *Ds* excision and reinsertion

markers, with *Agrobacterium* harbouring *iAc* constructs containing a visual marker, s*gfpS65T*, we have been able to regenerate stable *Ds* insertion lines at a frequency of ~5%, in addition to *iAc/Ds* double transformants (Upadhyaya et al. 2006). Mapped single-copy *Ds*/T-DNA launch pads, produced using these constructs are highly suitable for efficient chromosomal region-directed insertion mutagenesis.

It is presumed that public availability of *Tos17*, T-DNA, *Ac/Ds* and *En/Spm* insertion lines and their flanking sequences will enable rice geneticists to find at least one insertion in any rice gene and several alleles in most of the genes in the near future. This will greatly facilitate unravelling the functions of agronomically important rice genes.

Yet another type of insertional mutagenesis is activation tagging. In the classical activation tagging approach, random insertion of a CaMV 35S enhancer element in the rice genome can result in over-expression of native genes (or even dormant genes) in all cell types of the plant. Such increased gene expression can create mutations for essential and redundant genes that are either not present or have no phenotype in knockout collections. This gain of function approach produces dominant mutations affecting the transcriptional control of genes without altering the functional gene product. A sizable number of T-DNA activation tagging lines have been produced by research groups in France, Korea, China and Taiwan (see Guiderdoni et al. 2007). Further refinement of activation tagging comes from the development of extensive GAL4 enhancer trapping resources in rice which enable transgene expression to be targeted to specific cell types (see Johnson et al. 2007). In the first step of a two step process known as "transactivation" a large number of GAL4 enhancer trapping "driver" lines are generated and patterns of reporter gene expression are characterized. "responder" lines are then produced having transgenes of interest cloned downstream of the UAS element to which GAL4 binds. Crosses between driver and responder lines result in transactivation of the target genes by GAL4 showing the specific expression profile of each individual driver. In addition, random deployment of the UAS element into the rice genome, followed by crosses to specific driver lines, should enable activation tagging to be carried out in specific cell types of the plant. Cell type-specific activation tagging has the potential to uncover novel mutations that are missed or "averaged out" by the classical activation tagging technique (see Johnson et al. 2007).

11.9 From Rice to Other Cereals – Comparative Genomics

Moore et al. (1995) first showed by comparative mapping studies using RFLP markers that the organisation of chromosomes (or chromosome segments) of the various cultivated cereals are essentially syntenic to that of rice chromosomes. The so-called "Crop-Circle" model where the different cereal genomes can be aligned around a circle with the diameter of the circle representing the genome size has been updated and refined (Devos 2005). The synteny between rice and other cereals has been used extensively in map based cloning of genes in cereals such as wheat.

The small size of the rice genome allows markers from rice to be used in a cross species-specific way, often "jumping" repetitive regions of the wheat genome.

A significant continuation of rice genome sequencing is the BAC-end sequencing of wild-type species of *Oryza*, each with a coverage of 10 to 19% of their genome (http://www.omap.org; Ammiraju et al. 2006). This makes rice a unique resource amongst the eukaryotes, with two genomes fully sequenced and a dozen other related genomes physically mapped and partially sequenced (see Cooke et al. 2007).

Rice gene sequences can be compared with those of other cereals allowing more definitive annotation of coding sequences. Promoter elements conserved across species can also be identified by comparative genomics pointing to functionally important motifs. Comparative expression profiling studies of coding as well as non-coding, noncoding sequences (siRNA and miRNAs) are also facilitating new discoveries of biological significance.

11.10 Challenges and Prospects

International collaboration is essential to fully utilise the tools and resources currently being developed for rice research worldwide. A set of lines with mutations in each of the predicted 50,000 genes together with integrated databases containing all the relevant information about each rice gene is critical for assigning gene function. Current annotation of genes is far from perfect and is an iterative process. Other plant genome sequences including that of *Arabidopsis* can assist in further training of gene prediction programs to further improve the annotation. There is a need for unified usage of pseudomolecule terminologies by different curating groups. There should be an emphasis on manual curation with back-up wet science to produce the most accurate gene models. Deep sequencing data can be used to assist in gene annotation especially to determine the 3' and 5' ends of mRNAs and to identify splice variants.

Precise phenotyping of mutants with common descriptors of characters between laboratories is also important in assigning gene function. Phenotyping can be performed effectively through collaborations between laboratories with complementary expertise. Towards this goal an International Rice Functional Genomics Consortium (IRFGC) has been formed which should provide a common platform for information and resource sharing. In this way the goal of understanding the structure and function of each gene in the rice genome can be achieved.

The application of the outputs of the rice functional genomics efforts will be via the use of naturally occurring, agronomically-useful alleles of rice or other *Oryza* genomes. Molecular markers will allow the rapid integration of these alleles into breeding programs. Following the association of a phenotype to a gene, the level of, or pattern of expression of that gene can be altered to achieve the desired effect. This can be done by looking for mutants in the gene (e.g. TILLING) or by using transgenic methods to either reduce gene expression by RNAi technology, or to

overexpress the gene of interest using suitable promoters. These specifically altered lines can then be incorporated into breeding programs. These transgenic solutions may be used if public antipathy to GM crops can be overcome.

One challenge is identifying genes involved in complex traits of agronomic significance such as tolerance to abiotic stress. It is likely that there will be many genes with some effect and pinpointing critical genes will require inputs from all aspects of genetics and genomics. These characters will be of critical importance in altered environments caused by changing climate.

Finally, although rice is an important food crop in its own right, it is also a model for other cereals. Discoveries in rice can be applied to other cereal species such as maize, wheat and barley. With sequencing projects commencing in maize and wheat the functional genomics findings in rice will assist gene selection and breeding in these other cereals.

Acknowledgements The authors wish to thank Drs. Danny Llewellyn, Chris Helliwell for reviewing the manuscript and Drs. Andrew Eamens, Qian-Hao Zhu and Sadequr Rahman for critical reading of the manuscript.

References

Ammiraju JS, Yu Y, Luo M et al. (2005) Random sheared fosmid library as a new genomic tool to accelerate complete finishing of rice (*Oryza sativa* spp. Nipponbare) genome sequence: sequencing of gap-specific fosmid clones uncovers new euchromatic portions of the genome. Theor Appl Genet 111:1596–1607

Ammiraju JS, Luo M, Goicoechea JL et al. (2006) The *Oryza* bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus *Oryza*. Genome Res 16:140–147

An S, Park S, Jeong DH et al. (2003) Generation and analysis of end sequence database for T-DNA tagging lines in rice. Plant Physiol 133:2040–2047

An G, Lee S, Kim SH et al. (2005) Molecular genetics using T-DNA in rice. Plant Cell Physiol 46:14–22

Antonio BA, Buell R, Yamazaki Y et al. (2007) Informatics resources for rice functional genomics. In: Upadhyaya NM (ed) Rice functional genomics-challenges, progress and prospects. Springer, New York, pp 356–394

Bhat RS, Upadhyaya NM, Chaudhury A et al. (2007) Chemical and irradiation induced mutants and TILLING. In: Upadhyaya NM (ed) Rice functional genomics-challenges, progress and prospects. Springer, New York, pp 149–180

Borsani O, Zhu J, Verslues PE et al. (2005) Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. Cell 123:1279–1291

Brar DS, Khush GS (1997) Alien introgression in rice. Plant Mol Biol 35:35–47

Brenner S, Johnson M, Bridgham J et al. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol 18:630–634

Chen JJ, Rowley JD, Wang SM (2000) Generation of longer cDNA fragments from serial analysis of gene expression tags for gene identification. Proc Natl Acad Sci U S A 97:349–353

Chen S, Jin W, Wang M et al. (2003) Distribution and characterization of over 1000 T-DNA tags in rice genome. Plant J 36:105–113

Chin HG, Choe MS, Lee SH et al. (1999) Molecular analysis of rice plants harboring an *Ac/Ds* transposable element-mediated gene trapping system. Plant J 19:615–623

Cooke R, Piégu B, Panaud O et al. (2007) From rice to other cereals: comparative genomics. In: Upadhyaya NM (ed) Rice functional genomics-challenges, progress and prospects. Springer, New York, pp 429–479

Devos KM (2005) Updating the 'crop circle'. Curr Opin Plant Biol 8:155–162

Dunn WB, Bailey NJ, Johnson HE (2005) Measuring the metabolome: current analytical technologies. Analyst 130:606–625

Eubel H, Braun HP, Millar AH (2005) Blue-native PAGE in plants: a tool in analysis of protein-protein interactions. Plant Methods 1:11

Fujino K, Sekiguchi H, Kiguchi T (2005) Identification of an active transposon in intact rice plants. Mol Genet Genomics 273:150–157

Goff SA, Ricke D, Lan TH et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92–100

Gowda M, Jantasuriyarat C, Dean RA et al. (2004) Robust-LongSAGE (RL-SAGE): a substantially improved LongSAGE method for gene discovery and transcriptome analysis. Plant Physiol 134:890–897

Gowda M, Venu RC, Jia Y et al (2007a) Use of robust-long serial analysis of gene expression to identify novel fungal and plant genes involved in host-pathogen interactions. Methods Mol Biol 354:131–144

Gowda M, Venu RC, Li H et al (2007b) *Magnaporthe grisea* infection triggers RNA variation and antisense transcript expression in rice. Plant Physiol 144:524–533

Greco R, Ouwerkerk PB, Taal AJ et al (2004) Transcription and somatic transposition of the maize En*/Spm* transposon system in rice. Mol Genet Genomics 270:514–523

Griffiths-Jones S, Saini HK, van Dongen S et al. (2008) miRBase: tools for microRNA genomics. Nucleic Acids Res 36:D154–D158

Guiderdoni E, An G, Yu S-M et al. (2007) T-DNA insertion mutants as a resource for functional genomics. In: Upadhyaya NM (ed) Rice functional genomics-challenges, progress and prospects. Springer, New York, pp 182–321

Hirochika H (2001) Contribution of the *Tos17* retrotransposon to rice functional genomics. Curr Opin Plant Biol 4:118–122

Hirochika H, Sugimoto K, Otsuki Y et al (1996) Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci U S A 93:7783–7788

Hirochika H, Guiderdoni E, An G et al (2004) Rice mutant resources for gene discovery. Plant Mol Biol 54:325–334

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800

Iyer-Pascuzzi AS, Sweeny MT, Sarla N et al (2007) Use of naturally occurring alleles for crop improvement. In: Upadhyaya NM (ed) Rice functional genomics-challenges, progress and prospects. Springer, New York, pp 110–147

Jiang N, Bao Z, Zhang X et al (2003) An active DNA transposon family in rice. Nature 421:163–167

Johnson AAT, Yu S-M, Tester M (2007) Activation tagging systems in rice. In: Upadhyaya NM (ed) Rice functional genomics-challenges, progress and prospects. Springer, New York, pp 333–353

Jones-Rhoades MW, Bartel DP (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol Cell 14:787–799

Khan M, Takasaki H, Komatsu S (2005) Comprehensive phosphoproteome analysis in rice and identification of phosphoproteins responsive to different hormones/stresses. J Proteome Res 4:1592–1599

Khush GS (1997) Origin,dispersal, cultivation and variation of rice. Plant Mol Biol 35:25–34

Kikuchi S, Satoh K, Nagata T et al (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice. Science 301:376–379

Kikuchi S, Wang G-L, Li L (2007) Genome-wide RNA expression profiling in rice. In: Upadhyaya NM (ed) Rice functional genomics-challenges, progress and prospects. Springer, New York, pp 31–59

Komatsu M, Chujo A, Nagato Y et al (2003a) *FRIZZY PANICLE* is required to prevent the formation of axillary meristems and to establish floral meristem identity in rice spikelets. Development 130:3841–3850

Komatsu M, Shimamoto K, Kyozuka J (2003b) Two-step regulation and continuous retrotransposition of the rice LINE-type retrotransposon *Karma*. Plant Cell 15:1934–1944

Komatsu S (2007) Rice proteomics: a step toward functional analyses of the rice genome. In: Upadhyaya NM (ed) Rice functional genomics-challenges, progress and prospects. Springer, New York, pp 61–89

Komatsu S, Zang X, Tanaka N (2006) Comparison of two proteomics techniques used to identify proteins regulated by gibberellin in rice. J Proteome Res 5:270–276

Koornneef M, Dellaert LW, van der Veen JH (1982) EMS- and radiation-induced mutation frequencies at individual loci in *Arabidopsis thaliana* (L.) Heynh. Mutat Res 93:109–123

Lapidot M, Pilpel Y (2006) Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. EMBO Rep 7:1216–1222

Leung H, Wu C, Baraoidan M et al (2001) Deletion mutants for functional genomics: progress in phenotyping, sequence assignment, and database development. In: Khush G, Brar D, Hardy B (eds) Rice genetics IV. New Delhi Science Publishers, Inc., India

Li L, Wang X, Xia M et al (2005) Tiling microarray analysis of rice chromosome 10 to identify the transcriptome and relate its expression to chromosomal architecture. Genome Biol 6:R52

Li L, Wang X, Stolc V et al (2006) Genome-wide transcription analyses in rice using tiling microarrays. Nat Genet 38:124–129

Li X, Song Y, Century K et al (2001) A fast neutron deletion mutagenesis-based reverse genetics system for plants. Plant J 27:235–242

Lonhosky PM, Zhang X, Honavar VG et al (2004) A proteomic analysis of maize chloroplast biogenesis. Plant Physiol 134:560–574

Lu C, Jeong DH, Kulkarni K et al (2008) Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). Proc Natl Acad Sci U S A 105:4951–4956

Ma L, Chen C, Liu X et al (2005) A microarray analysis of the rice transcriptome and its comparison to *Arabidopsis*. Genome Res 15:1274–1283

Matsumoto T, Wing RA, Han B et al (2007) Rice genome sequence: the foundation for understanding the genetic systems. In: Upadhyaya NM (ed) Rice functional genomics-challenges, progress and prospects. Springer, New York, pp 5–30

Matsumura H, Reich S, Ito A et al (2003) Gene expression analysis of plant host-pathogen interactions by SuperSAGE. Proc Natl Acad Sci U S A 100:15718–15723

McCouch SR, Kochert G (1988) Molecular mapping of rice chromosomes. Theor Appl Genet 76:815–829

Meyers BC, Tej SS, Vu TH et al (2004a) The use of MPSS for whole-genome transcriptional analysis in Arabidopsis. Genome Res 14:1641–1653

Meyers BC, Vu TH, Tej SS et al (2004b) Analysis of the transcriptional complexity of *Arabidopsis thaliana* by massively parallel signature sequencing. Nat Biotechnol 22:1006–1011

Mizuno H, Ito K, Wu J et al (2006a) Identification and mapping of expressed genes, simple sequence repeats and transposable elements in centromeric regions of rice chromosomes. DNA Res 13:267–274

Mizuno H, Wu J, Kanamori H et al (2006b) Sequencing and characterization of telomere and subtelomere regions on rice chromosomes 1S, 2S, 2L, 6L, 7S, 7L and 8S. Plant J 46:206–217

Moore G, Devos KM, Wang Z et al (1995) Cereal genome evolution. Grasses, line up and form a circle. Curr Biol 5:737–739

Morin RD, Aksay G, Dolgosheina E et al (2008) Comparative analysis of the small RNA transcriptomes of *Pinus contorta* and *Oryza sativa*. Genome Res 18:571–584

Nobuta K, Venu RC, Lu C et al (2007) An expression atlas of rice mRNAs and small RNAs. Nat Biotechnol 25:473–477

O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. J Biol Chem 250:4007–4021

Patankar S, Munasinghe A, Shoaibi A et al (2001) Serial analysis of gene expression in *Plasmodium falciparum* reveals the global expression profile of erythrocytic stages and the presence of anti-sense transcripts in the malarial parasite. Mol Biol Cell 12:3114–3125

Phillips JR, Dalmay T, Bartels D (2007) The role of small RNAs in abiotic stress. FEBS Lett 581:3592–3597

Raghavan C, Naredo E, Wang H et al (2007) Rapid method for detecting SNPs on agarose gels and its application in candidate gene mapping. Mol Breed 19:87–101

Ronaghi M, Karamohamed S, Pettersson B et al (1996) Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem 242:84–89

Saha S, Sparks AB, Rago C et al (2002) Using the transcriptome to annotate the genome. Nat Biotechnol 20:508–512

Sallaud C, Gay C, Larmande P et al (2004) High throughput T-DNA insertion mutagenesis in rice: a first step towards in silico reverse genetics. Plant J 39:450–464

Sato S, Soga T, Nishioka T et al (2004) Simultaneous determination of the main metabolites in rice leaves using capillary electrophoresis mass spectrometry and capillary electrophoresis diode array detection. Plant J 40:151–163

Satoh K, Doi K, Nagata T et al (2007) Gene organization in rice revealed by full-length cDNA mapping and gene expression analysis through microarray. PLoS ONE 2:e1235

Shimono M, Sugano S, Nakayama A et al (2007) Rice WRKY45 plays a crucial role in benzothiadiazole-inducible blast resistance. Plant Cell 19:2064–2076

Stolc V, Li L, Wang X et al (2005) A pilot study of transcription unit analysis in rice using oligonucleotide tiling-path microarray. Plant Mol Biol 59:137–149

Sunkar R, Chinnusamy V, Zhu J et al (2007) Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. Trends Plant Sci 12:301–309

Sunkar R, Zhou X, Zheng Y et al (2008) Identification of novel and candidate miRNAs in rice by high throughput sequencing. BMC Plant Biol 8:25

Suzuki T, Eiguchi M, Satoh H et al (2005) A modified TILLING system for rice mutant screening. Rice Genet News Lett 22:89–91

Takahashi H, Hotta Y, Hayashi M et al (2005) High throughput metabolome and proteome analysis of transgenic rice plants (*Oryza sativa* L.). Plant Biotechnol 22:47–60

Tanaka N, Mitsui S, Nobori H et al (2005) Expression and function of proteins during development of the basal region in rice seedlings. Mol Cell Proteomics 4:796–808

Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277:1063–1066

Tarpley L, Roessner U (2007) Metabolomics: enabling sytstems-level phenotyping in rice functional genomics. In: Upadhyaya NM (ed) Rice functional genomics-challenges, progress and prospects. Springer, New York, pp 91–107

Tarpley L, Duran AL, Kebrom TH et al (2005) Biomarker metabolites capturing the metabolite variance present in a rice plant developmental period. BMC Plant Biol 5:8

Till BJ, Cooper J, Tai TH et al (2007) Discovery of chemically induced mutations in rice by TILLING. BMC Plant Biol 7:19

Upadhyaya NM, Zhu QH, Zhou XR et al (2006) *Dissociation* (*Ds*) constructs, mapped *Ds* launch pads and a transiently-expressed transposase system suitable for localized insertional mutagenesis in rice. Theor Appl Genet 112:1326–1341

Velculescu VE, Zhang L, Vogelstein B et al (1995) Serial analysis of gene expression. Science 270:484–487

Wing RA, Kim H, Goicoechea JL et al (2007) The *Oryza* Map Alignment Project (OMAP): a new resource for comparative genome studies within *Oryza*. In: Upadhyaya NM (ed) Rice functional genomics-challenges, progress and prospects. Springer, New York, pp 395–409

Xu Y, Zhou X, Zhang W (2008) MicroRNA prediction with a novel ranking algorithm based on random walks. Bioinformatics 24:i50–58

Yammamoto T, Lin H, Sasaki T et al (2000) Identification of heading date quantitative trait locus *Hd6* and characterization of its epistatic interactions with *Hd2* in rice using advanced backcross progeny. Genetics 154

Yano M (2001) Genetic and molecular dissection of naturally occurring variation. Curr Opin Plant Biol 4:130–135

Yazaki J, Shimatani Z, Hashimoto A et al (2004) Transcriptional profiling of genes responsive to abscisic acid and gibberellin in rice: phenotyping and comparative analysis between rice and *Arabidopsis*. Physiol Genomics 17:87–100

Yu J, Hu S, Wang J et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296:79–92

Yu J, Wang J, Lin W et al (2005) The Genomes of *Oryza sativa*: a history of duplications. PLoS Biol 3:e38

Zhao W, Wang J, He X et al (2004) BGI-RIS: an integrated information resource and comparative analysis workbench for rice genomics. Nucleic Acids Res 32:D377–D382

Zhu QH, Hoque MS, Dennis ES et al (2003) *Ds* tagging of *BRANCHED FLORETLESS 1* (*BFL1*) that mediates the transition from spikelet to floret meristem in rice (*Oryza sativa* L). BMC Plant Biol 3:6

Zhu Q-H, Eun MY, Han C-D et al (2007) Transposon insertional mutants: a resource for rice functional genomics. In: Upadhyaya NM (ed) Rice functional genomics-challenges, progress and prospects. Springer, New York, pp 223–271

Zhu Q-H, Spriggs A, Matthew L et al (2008) A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. Genome Res. doi:10.1101/gr.075572.107

Chapter 12
Genomics for Wheat Improvement

Michael G. Francki

Abstract The ability to meet the demands of global food production will require efficient means to develop modern cultivars adaptable to a range of adverse environmental conditions in marginal wheat production zones. Breeding programs will be relying on the tools used to track allelic combinations contributing to trait variation through DNA marker-assisted selection and efficient selection of genotypes expressing desirable phenotypes in target environments. The recent developments in wheat genomics have provided resources to develop new molecular markers and strategies for genetic analysis and identification of marker-trait associations. Included are new DNA marker technologies capable of developing high resolution genetic maps and QTL mapping allowing detection of trait variation at specific loci. The ability to locate the chromosomal region associated with phenotypic variation provides a leading edge towards developing functionally-associated markers (FAM) to track alleles in a breeding program. Map-based cloning, comparative genomics and sequencing the wheat genome provides current and future opportunities for discovering genes responsible for trait variation. Determining the function of newly discovered genes will allow their effective use in wheat improvement as FAM markers for marker-assisted breeding. Therefore, transgenic plants overexpressing or silencing genes by RNA interference (RNAi) and non-transgenic approaches such as virus-induced gene silencing (VIGS) and Targeting Induced Local Lesions IN Genomes (TILLING) provide strategies to determine gene function and their effects on phenotypic variation. Transgenic wheat plants and TILLING approaches also has the advantage in developing potential new varieties but the latter would be the only option in countries where the release of genetically modified wheat is constrained.

M.G. Francki
Department of Agriculture and Food Western Australia, State Agricultural Biotechnology Centre, Murdoch University, Western Australia and Molecular Plant Breeding Cooperative Research Centre
e-mail: michael.francki@agric.wa.gov.au

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_12,

12.1 Introduction

Bread wheat (*Triticum aestivum* L.) is one of the most staple food crops and is the most important carbohydrate source for human consumption. The major grain protein is gluten, an elastic form of protein that gives rise to unique viscoelastic properties of dough produced from bread wheat flour suitable for a range of end-products. Global wheat production averages 216 million hectares and leads all crops in total production area, including rice and maize. Since, the 1950s, there has been substantial increase in wheat production, largely due to genetically improved higher yielding, disease resistant, semi-dwarf cultivars. However, since the early 1990's wheat production has remained relatively constant with over 500 million tonnes per annum and with the imminent global population increase, projected requirements will be in excess of 750 million tonnes annually (Lantican et al. 2002). To meet the future global demands for grain, high-yielding modern cultivars will need to be achieved either through increasing production area, increasing yield per sown area or both. It is likely that wheat production will rely more on developing new higher-yielding cultivars in marginal environments. The existing and future threats of global warming and climate change will no doubt place further pressure to increase genetic gain and adaptability of new varieties to meet the supply of the growing food demands. Therefore, wheat breeding will play a primary role in developing modern cultivars that are adapted to current and future adverse environments. However, developing modern wheat cultivars to maintain or increase yields either in existing or marginal wheat production areas will require continuous innovation and adoption of new breeding methodologies and technologies.

Future increases in genetic gain will rely on building the knowledge-base and understanding of trait variation that enable wheat varieties to adapt to a range of environments. Although consistent expression of a desirable phenotypic in a target environment is the primary outcome, equally important is knowledge of the genetic, phyisological and biochemical interactions underpinning the control and environmental effects on trait variation. It is the collective interpretation of these interactions and how they can be manipulated that will drive selection of high-yielding modern wheat cultivars adaptable to a range of marginal environments. In the last 15 years, there has an exponential increase in a vast assortment of genomic resources for wheat and other related grasses, adding considerable value in increasing our knowledge of gene interactions that control phenotypic variation. Although wheat is the most important food source for the world, its genome composition makes it one of the most difficult to apply functional and comparative genomics for cultivar improvement. This chapter summarises recent advances in wheat genomics that offer potential for wheat improvement. The latest developments in genetic analysis, gene discovery and gene function provide non-transgenic and transgenic options for breeding programs to adopt new tools and strategies for developing high-yielding cultivars adaptable to changing environments and marginal production zones suitable for high-income and developing countries.

12.2 Genetic Analysis

Wheat is one of the complex genetic systems of the major crop species with an allohexaploid genome (2n = 6x = 42, genome AABBDD) of approximately 16,000 megabases in size and 1-C value of 16.5–19.5 pg (reviewed in Appels et al. 2003). It originated from the hybridization of allotretraploid emmer wheat (*T. turgidum* ssp. *Dicoccum* 2n = 4x = 28, genome AABB) with diploid *Aegilops tauschii* (2n = 2x = 14, genome DD) (McFadden and Sears 1946). The large size, complex arrangement of repetitive sequences and the hexaploid nature of the wheat genome (reviewed in Appels et al. 2003), makes the development of genomic resources and application to breeding programs a challenging task. Nevertheless, the global economic importance of wheat and the need to develop higher yielding wheat varieties has seen the development of latest genomic tools and technologies to understand the genetic control of a range of morphological characteristics, grain quality and tolerances to biotic and abiotic stresses for adaptation. Traits are either quantitatively or qualitatively inherited and traditional strategies have heavily relied on selecting the desired phenotype for the target environment throughout the breeding process. Selection based on DNA or functionally-associated markers linked to genes controlling trait variation are replacing the often arduous and time consuming phenotypic evaluation in early generation selection to predict individuals with specific alleles that are likely to express a particular characteristic in a target environment. The development of genetic maps and subsequent QTL and linkage disequilibrium (LD) mapping is a prerequisite to identify DNA markers linked to genes controlling qualitative and quantitative traits prior to implementation in marker-assisted selection. The abundance in genomic resources for wheat have been exploited to build comprehensive genetic maps for QTL and LD mapping to identify DNA markers linked to important genes for use in early generation selection during wheat improvement.

12.2.1 Genetic Maps of Wheat

Wheat genetic maps were first developed using restriction fragment length polymorphic (RFLP) markers (Chao et al. 1989; Devos et al. 1992, 1993, 1995; Xie et al. 1993; Jia et al. 1996; Devos and Gale 1997) and have recently updated with PCR based markers. Although amplified fragment length polymorphic (AFLP) markers have contributed to increasing the resolution of genetic maps (Vos et al. 1995), abundant and locus-specific microsatellite or simple sequence repeat (SSR) markers (Roder et al. 1998; Pestsova et al. 2000; Gupta et al. 2002) are most favoured for genetic mapping in wheat. The International Triticeae Mapping Initiative (ITMI) population derived from cross between synthetic and domesticated parents (W-7984/Opata-85) is available to the research community and provides a valuable resource to determine the chromosomal location of newly developed

markers and their location with respect to other wheat molecular markers (http://wheat.pw.usda.gov/ITMI/). Although the ITMI population is of considerable use for wheat genetic mapping and identifying genes important for wheat breeding, the parents and population have limited use to practical wheat breeding. Therefore, numerous genetic maps based on breeders homozygous or heterozygous populations (doubled haploid, recombinant inbred, F2 and backcross populations) have been developed by wheat improvement programs world-wide to identify genes and QTL for specific traits of agronomic and regional importance. An example of constructing genetic maps from breeders populations using RFLP, AFLP and SSR markers are described in Chalmers et al. (2001). The maps have been used for the genetic analysis of a range of quality, biotic, abiotic and morphological traits important for Australian breeding programs (see separate articles in special issue of Australian Journal Agricultural Research, Vol 52). However, the distribution of markers across individual maps has made the compilation of markers for any given linkage group unwieldy. In a more recent study, approximately 1,2002004 markers from four independent genetic maps were merged into a single consensus map (Somers et al.). The consensus map is a particularly useful reference point for targeting additional markers for specific chromosome regions for fine mapping of QTL without the need to cross-reference independent maps with common markers.

The US Wheat Genome Project, funded by the National Science Foundation, developed an extensive core collection of expressed sequence tags (ESTs). As of March 2008, there were in excess of 1,000,000 ESTs represented in public domain databases (http://www.ncbi.nlm.nih.gov/) from *T. aestivum*, *T. monococcum*, *Aegilops speltoides* and *T. turgidum* (Chao et al. 2006; http://wheat.pw.usda.gov/NSF/). The EST collection was generated from at least 42 cDNA libraries, representing a comprehensive set of sequences expressed at various stages of plant development and under certain biotic and abiotic stress conditions. In order to identify the gene rich regions in the wheat genome, a selection of 7600 ESTs were used as RFLP probes to assign at least 16,000 loci to specific chromosomal regions by deletion bin mapping (Zhang et al. 2004; Lazo et al. 2004; Qi et al. 2003, 2004). An evaluation of 3025 ESTs identified 29% of the wheat genome having major and minor gene-rich regions whereas 30% of the wheat genes are in recombination-poor regions (Erayman et al. 2004).

Despite generating a large collection and locating expressed portion of the genome to specific regions of chromosomes, ESTs are also a valuable source to add value to genetic map development and identifying marker-trait associations. Searching for short repetitive elements within databases identified in excess of 1,0002002 wheat ESTs containing SSRs (Eujayl et al. ; Kantety et al. 2002) and can be used for genetic map construction. However, the level of EST-SSR polymorphism across bread wheat accessions is lower than those for genomic derived SSRs, implying alterations in open reading frames have functional constraints in cultivated wheat (Gupta et al. 2003; Gao et al. 2004; Gadaleta et al. 2007). Nevertheless, EST-SSR polymorphic between W-7984 and Opata-85 were mapped in the ITMI reference genetic map and showed a non-random distribution across homoeologous genomes with a high degree of clustering at the distal ends of chromosomes.

This indicated that EST-SSR were confined to the genic regions of the wheat genome and in areas of high recombination (Gao et al. 2004; Yu et al. 2004). Not surprisingly, the EST-SSR markers were identified in similar regions of wheat chromosomes when ESTs were used in RFLP analysis (Zhang et al. 2004; Lazo et al. 2004; Qi et al. 2003, 2004). Apart from increasing the resolution of genetic maps with higher marker density, EST-SSR markers are of particular interest since they represent the transcribed region of the wheat genome. A study by Gao et al. (2004) identified that 74% of the mapped EST-SSR showed significant similarities to known genes which may be used to associate putative function with trait variation.

SSR (genomic and EST-derived) remain the popular choice for genetic map construction, marker-trait association and selection of breeding germplasm. Therefore, recent efforts have been made in developing more robust platforms supporting high throughput genotyping of wheat germplasm using SSR markers. Multiplex-ready PCR has been developed as an automated genotyping technology using standardized protocols with the flexibility in fluorescence labelling of SSRs and multiplexing capability (Hayden et al. 2008a, b). The high throughput capability of multiplex ready PCR has recently been realised with the deployment of three QTL for flag leaf and glume resistance to *Stagonospora nodorum* (Uphaus et al. 2007; Shankar et al. 2008) across selected Australian germplasm. The selection from 2000 backcross individuals using flanking SSR markers screened by Multiplex PCR technology, totalled in excess of 14,000 datapoints generated in less than 1 month (M. Francki, unpublished data).

Transcribed sequences remain attractive DNA markers because assumptions on gene function can be associated with specific traits of interest. An alternative marker generated from ESTs includes single nucleotide polymorphism (SNP). However, SNP analysis in hexaploid wheat is more challenging than comparable studies in diploids, requiring the development of genome-specific SNP markers to ensure one of the genomes is amplified to reveal haplotypes in the specific genome being studied. Several studies used EST contig assembly to identify SNPs (Mochida et al. 2004; Ravel et al. 2006; Somers et al. 2003) and distinguish polymorphism between wheat cultivars (for example, Somers et al. 2003; Ravel et al. 2006; Wang et al. 2007). As yet, there are no abundant SNP markers for constructing high-resolution genetic maps. However, improvements in efficient SNP discovery technologies, similar to the Multiplex-ready PCR for automated SSR analysis, coupled with low cost marker platforms is likely increase the development of high resolution SNP-based genetic maps of wheat and suitable marker systems for implementation in breeding programs.

12.2.2 Quantitative Trait Loci and Linkage Disequilibrium

Many traits of agronomic significance show continuous or quantitative variation controlled by several genes within chromosomal regions having additive or dominance effects are known as quantitative trait loci (QTL). QTL are detected by statistical

association between marker alleles on a genetic linkage map and individual trait differences from a biparental mapping population, defining the chromosomal region containing alleles controlling phenotypic variation (for review, see Remington and Purugganan 2003). Despite the majority of most studies using structured bi-parental mapping populations, QTL may represent only a small part of the genetic architecture of a particular trait. Identical-by-descent (IBD) provides a powerful statistical method that takes into consideration alleles that are common by descent contributing to trait variation in complex populations used in pedigrees of arbitrary size and complexity (Almasy and Blangero 1998; Crepieux et al. 2004). There are now many reported QTL studies for a comprehensive range of biotic, abiotic, morphological and grain quality traits in wheat. The associated markers in wheat have been useful in defining the effects of gene interactions for trait variation and tracking allelic combinations in breeding for wheat improvement. However, the precision of QTL mapping is limited to the DNA polymorphism between parents of biparental populations, the size of the population, the degree and accuracy of phenotyping and distribution of chiasma across the genome. Therefore, to complement marker-trait associations, a non-random association of alleles between linked or unlinked loci is estimated (linkage disequilibrium, LD) based on a large collection of accessions unrelated by descent (for review, see Gupta et al. 2005). Thus, association mapping is one of the uses of LD. However, a prerequisite in applying LD is to estimate the genetic distance over which LD will decay back to random associations of alleles, facilitating the prediction of marker density required to effectively associate genotypes with traits (Rafalski and Morgante 2004). A recent study by Somers et al. (2007) has shown that bread and durum wheat populations could resolve marker-trait associations within a genetic distance of 5 cM. However, it should be noted that LD decays with genetic distance varies, as has been previously noted between populations of barley land races and elite cultivars (Caldwell et al. 2006), therefore an appropriate population structure is an important consideration prior to development of LD studies (MacKay and Powell 2007). In a recent study by Breseghello and Sorrells (2006), the same markers identified in QTL analysis and association mapping studies showed significant linkage with kernel characteristics, providing complementary analysis of marker-trait associations prior to implementation of markers for wheat breeding. The full potential of the effects of LD will be realised when further association mapping complements QTL studies for a wide range of traits in wheat.

12.3 Wheat Gene Discovery

High-resolution genetic maps and refinements in QTL analysis have made significant contributions to understanding simple and complex genetic traits. However, an intriguing aspect is the identification of gene(s) within the marker interval that control trait expression. The increasing genomic resources available for wheat enable the isolation of genes within QTL and the development of diagnostic markers for

tracking allelic variation in breeding programs. Bacterial artificial chromosome (BAC) libraries and map-based cloning, comparative genomics and advances in sequencing the wheat genome provide the necessary resources to isolate and characterise genes that control trait variation.

12.3.1 Wheat BAC Libraries and Map-based Cloning

The cloning of large genomic fragments of the wheat genome and ordering BACs into contiguous segments provides physical maps to align with genetic recombination maps. However, one of the major limitations in hexaploid wheat is its large size and complexity of repetitive DNA making the construction of BAC libraries and subsequent contig assembly a challenging task. Nevertheless, several reports have developed BAC libraries of the hexaploid wheat genome with average insert sizes of 0.79–140 with 3.1–5.6x wheat genome equivalents (Allouis et al. 2003; Liu et al. 2000; Ling and Chen 2005; Nilmalgoda et al. 2003). In order to assist BAC contig assembly, sub-genome BAC libraries of the A genome from *T. monococcum* (Lijavetzky et al. 1999), D genome from *T. tauschii* (Moullet et al. 1999) and AB genome from *T. turgidum* var. *durum* (Cenci et al. 2003) will simplifying physical map assembly to specific genomes. Furthermore, BAC library construction from flow sorted wheat chromosomes of 1B (Janda et al. 2006); 1D, 4D, 6D (Janda et al. 2004) and 3B (Safar et al. 2004) and similar resources for all wheat chromosomes will accelerate future development of sequence-ready physical contig maps of the wheat genome. It is interesting to note that BAC coverage ranged from 6–14.1x chromosome coverage, reflecting an increased probability of isolating a gene of interest using chromosome specific rather than whole genome BAC libraries. Nevertheless, both libraries are complementary resources for the isolation of genes using for map-based cloning approaches.

The development of genetic maps with high-density molecular has facilitated the isolation of BAC clones within a particular marker interval containing the gene of interest. High throughput fingerprinting such as those described by Luo et al. (2003), followed by contig assembly and anchoring using molecular markers provides the alignment of physical and genetic maps to the target region. Subsequent sequencing of individual BAC clones and annotation of repetitive and low-copy sequences provides information for candidate gene identification. There are now several genes of agronomic importance for wheat breeding that have been isolated using BAC or other large-insert libraries and principles of map-based cloning. These include rust resistance genes *Lr1* (Cloutier et al. 2007), *Lr10* (Feuillet et al. 2003), *Lr21* (Huang et al. 2003), powdery mildew resistance (*Pm3b*) (Yahiaoui et al. 2004), vernalisation genes *Vrn1* and *Vrn2* (Yan et al. 2003, 2004). In each case, the isolation of these genes and their contribution to the phenotypic effect was confirmed using functional studies. Although only seven wheat genes have been cloned in recent years, the availability of BAC libraries and the increasing efforts in developing high-resolution

genetic maps of target chromosomal regions will likely see an increase in the number of other agronomically important genes isolated using map-based approaches in the near future.

12.3.2 Comparative Genomics with Model Plant Species and Grasses

An alternative approach to identifying genes controlling trait variation in wheat is to compare genes from closely or distantly related species. The complete genome sequence from the model plant, *Arabidopsis thaliana* (The *Arabidopsis* Genome Initiative 2000) and rice (Goff et al. 2002; Yu et al. 2002; International Rice Genome Sequencing Project 2005) has enabled the large scale comparison of gene and genomes between model and crops plants. The identification of duplicated genes within species (paralogs) and similar genes across species sharing a common progenitor (orthologs) provide the opportunity to compare and contrast model and crop genomes that evolved through evolution (for review see Francki and Appels 2007). Wheat gene orthologs that may have similar function can be identified by comparative sequence and gene structure analysis from taxonomically distant species such as Arabidopsis that diverged more than 200 million years ago (MYA) (Kellog 2001). However, wheat and rice are members of the Pooideae family that diverged from a common ancestral genome about 70 MYA (Wolfe et al. 1989) allowing rice to be an ideal species to bridge the evolutionary gap between monocots and dicots for comparative genomic analysis. The studies by Mullan et al. (2007) provide an example of tracing the evolutionary relationship of gene orthologs between Arabidopsis, rice and wheat. Arabidopsis genes with known function in Na^+ transport showed similar intron-exon boundaries in rice and wheat despite the effects of over 200 million years of evolutionary forces since divergence of monocots and dicots. However, it must be noted that comparison of gene structure across divergent species such as wheat and Arabidopsis does not necessarily imply similar gene function and studies are needed to determine the functional role of orthologs on trait variation in each species. In some instances, subtle differences in gene orthologs, even across closely related species, can account for major differences in gene function and trait expression. A study by Francki et al. (2006) compared invertase genes of rice with counterparts in wheat and perennial ryegrass (estimated divergence of wheat and perennial rye grass from a common ancestor 15–30 MYA (Soreng and Davis 1998). Unlike wheat and perennial ryegrass, rice does not accumulate fructan and is therefore, devoid of fructosyltransferase activity. However, rice invertase genes share similar intron-exon structure as functional fructosyltransferase genes from perennial ryegrass, indicating subtle changes in a common gene progenitor caused significant phenotypic differences during evolution. Not surprisingly, therefore, the function of wheat orthologs having either invertase or fructosyltransferase activity was unpredictable based on comparative genomic analysis.

QTL mapping provides a direct means to account for trait variation at a specific locus by which physical maps can be aligned and putative candidate genes identified from annotated genome sequence. Therefore, the assumption is that large scale conservation in gene content and order (macrosynteny) between wheat and other grasses genomes (Gale and Devos 1998; reviewed in Appels et al. 2003) has been a popular approach for cross-species transfer of information for candidate gene identification. However, it is apparent that microsyntenic rearrangements in gene content and order is common between genomes (Francki et al. 2004; La Rota and Sorrells 2004; Li et al. 2004; Munkvold et al. 2004; Miftahudin et al. 2004; Randhawa et al. 2004; Singh et al. 2007; Sorrells et al. 2003). Therefore, the mosaic of conserved and non-conserved regions across genomes can either benefit or complicate the use of rice as a model for cross-species transfer of information. Studies identifying candidate genes controlling QTL for resistance to Fusarium head blight (FHB) highlight successes and failures in the application of comparative genomics for candidate gene identification. For example, a QTL for resistance to FHB on chromosome 2D has identified a multi-drug resistant associated protein effective against mycotoxin accumulation on a syntenic region on rice chromosome 4 (Handa et al. 2008). However, a QTL for resistance on chromosome 3B failed to identify a candidate gene on a syntenic region on rice chromosome 1 (Liu and Anderson 2003) but identified a sequence with similarity to a protein kinase domain characteristic of disease resistance genes on rice chromosome 11 (Shen et al. 2006).

The complications caused by micro-rearrangements have prompted an alternative model species be proposed for comparative genomics in wheat. The small genome size (~400 Mbp) and closer taxonomic relationship to members of the Triticeae with an estimated divergence from a common progenior of 35–40 MYA (Bossolini et al. 2007) identified *Brachypodium sylvaticum* and *B. distachyon* as potential model species for comparative genomics (Draper et al. 2001). The recent development of a BAC library of Brachypodium will form the basis for comparing gene and genome organisation across the grasses (Foote et al. 2004). However, initial studies indicate that Triticeae-Brachypodium microsyteny is quantitatively similar to Triticeae-rice and, therefore, unlikely to substitute for the large scale analysis of Triticeae genomes (Bossolini et al. 2007). Nevertheless, Brachypodium offers potential as a model species to complement the wheat-rice comparative analysis for wheat gene discovery.

12.3.3 Wheat Genome Sequencing

Whole genome sequence of rice (Goff et al. 2002; Yu et al. 2002) or clone-by-clone sequencing of the rice (International Rice Genome Sequencing Project 2005) and Arabidopsis (The Arabidopsis Genome Initiative 2000) genomes has demonstrated that most recent technologies are capable of sequencing whole plant genomes. However, the wheat genome is 30x and 128x the size of the rice and Arabidopsis genomes, respectively, and contains >70% repetitive DNA making sequencing

and annotation of the entire genome an arduous task. Nevertheless, the genomic resources available in wheat including the most comprehensive plant EST collection, high density genetic maps, BAC libraries and extensive genetic analysis provides the basis to assemble physical maps and annotate the wheat genome sequence to identify the genes controlling trait variation. Despite the resources available, whole genome sequencing of wheat will not be completed in the foreseeable future unless developments in sequencing technologies can be achieved at higher throughput and lower cost than is currently available (Stein 2007). It is unlikely that shot-gun sequencing would be the strategy of choice as the complexity of the genome will preclude accurate sequence assembly and annotation. A pilot study to determine the feasibility of sequencing the gene rich regions estimated that only one third of the genome would need to be sequenced using a clone-by-clone approach (Erayman et al. 2004). However, more recent investigations of randomly sequenced wheat BACs showed a gene density of 1 gene per 75 kb, indicating that genes are likely to be more evenly distributed across the genome than previously determined, substantiating the need for large-scale BAC sequencing of the wheat genome (Devos et al. 2005).

Sequencing the largest crop genome would require a consolidated effort by the international wheat research community. The International Wheat Genome Sequencing Consortium (IWGSC) is governed by six co-chairs representing countries including Australia, USA, France Switzerland and Japan (http://www.wheatgenome.org). The consortium is responsible for a strategic plan, coordinating project activities and ensuring sequence of the wheat genome and the DNA based tools are made available to the scientific community without restriction. Projects commissioned through the IWGSC include physical mapping, sequencing, structure and function and the molecular analysis of resistance loci on homoeologous group 3 chromosomes and wheat-Brachypodium-rice comparative analysis has commenced (http://www.wheatgenome.org). There is little doubt that completing the genome sequence of hexaploid wheat will require a highly coordinated, multidisciplinary and cross-institutional effort involving international consortium partners to assemble, annotate and determine the functional significance of wheat genes and repetitive sequences.

12.4 Gene Function

One of the critical facets of wheat gene discovery is determining functional roles in trait variation. Generally, this has been the rate-limiting step in identifying which genes can be manipulated to control expression of desirable traits for wheat improvement. A number of strategies can be achieved to decipher functional attributes of wheat genes. These include overexpression and RNA interference (RNAi), Targeting Induced Local Lesions in Genomes (TILLING) and virus induced gene silencing (VIGS). In some instances, strategies deployed to determine the functional role of genes can concurrently enhance germplasm for developing modern wheat varieties.

12.4.1 Transgenics and Overexpression

Transgenic plants up-regulating gene expression and their phenotypic differences to non-transgenic provide a means to determine gene function. Some studies have used Arabidopsis and tobacco as preferred species for determining gene function by transgenic analysis, as they are more amenable to transformation than wheat. Table 12.1 summarises wheat genes that have been transformed in taxonomically related or unrelated species and their phenotypic effect in order to elucidate gene function. In some instances, over-expression of wheat genes has produced the desired phenotypic effect but the precise function of the gene is unknown. For example, over-expression of wheat genes *Wcor15*, *Wcs19* and *WCOR410* in tobacco, Arabidopsis and strawberry, respectively, contribute to freezing tolerance but their direct functional role in biochemical, developmental or physiological pathways remain unclear (Table 12.1). In other instances, wheat genes have same function when overexpressed in a range of plant species. Overexpression of the wheat gene, *TaOxO*, produced high enzymic levels of oxalate oxidase providing broad-spectrum disease resistance in poplar, maize, sunflower, potato and soybean (Table 12.1). Although wheat genes expressed in other species provides information on gene function, their interactions with endogenous genes when over-expressed in wheat may not necessarily produce the desired phenotype.

The rapid advancements in gene discovery in the past 5 years have seen parallel increases in transgenic wheat expressing endogenous genes. Apart from determining gene function, transgenic wheat with modified gene expression has the potential in delivering new commercial varieties improved grain quality, disease resistance, abiotic stress tolerance and phenology. However, wheat is more difficult to transform than other major crop species relying on either biolistics or *Agrobacterium*-mediated methods to develop transgenic wheat with stably integrated genes. The advantages and disadvantages of each method have been recently reviewed (Shewry and Jones 2005; Bhalla 2006; Bhalla et al. 2006).

The major impact of transgenic wheat studies to date has been the modification of gene expression to alter grain quality characteristics for processing and end-use quality. The overexpression of endogenous genes for high molecular weight glutenin-subunits (HMW-GS) or allelic variants (Barro et al. 1997; Alvarez et al. 2000) have both positive and negative effects on dough characteristics. For example, overexpression of subunit classes *1Ax1* can have beneficial effects to breadmaking qualities whereas *1Dx5* showed detrimental effects producing ‘over-strong’ dough (Darlington et al. 2003; Blechl et al. 2007; Rooke et al. 1999; He et al. 2005; Barro et al. 1997). Therefore, the presence of an additional cysteine in *1Dx5* was responsible for a more highly cross-linked glutenin network preventing full hydration during dough mixing (Darlington et al. 2003), confirmed by over-expression of a truncated form of *1Dx5* (He et al. 2005). Similarly, overexpression of LMW-GS showed increase dough strength in durum wheat, opening possibilities for different end-products in niche markets (Tosi et al. 2004, 2005). On the contrary, overexpression of LMW-GS in bread wheat has negative effects on gluten visco-elastic properties (Masci et al. 2003), thereby affecting wheat end-use properties.

Table 12.1 Functional characterization of wheat genes by transgenic overexpression in other plant species and analysis of phenotypic effects

Wheat gene	Species transformed	Gene function	Phenotypic effect	Reference
TaERF1	Tobacco (*Nicotiana. tobacum* L.)	Interaction with protein kinase	Enhanced tolerance to cold, drought, salt	Xu et al. (2007a)
Wknox1a, Wknox1b, Wknox1d	Tobacco (*Nicotiana. tobacum* L.)			Morimoto et al. (2005)
Glu-1Dx5 and *Glu-1Dy10*	Rye (*Secale cereale* L.)	Gluten extensibility and elasticity	Increased polymerized glutelin fraction	Altpeter et al. (2004)
Glu-1Dx5	Maize (*Zea mays* L.)	Gluten extensibility and elasticity	Low pollen transmission efficiency	Sangtong et al. (2002); Scott et al. (2007)
TaOxO	Poplar (*Populus x euranericana*)	Oxalate oxidase activity	Partial resistance to *Septoria musiva*	Liang et al. (2001)
	Maize (*Zea mays* L.)	Oxalate oxidase activity	Resistance to european corn borer	Ramputh et al. (2002)
	Sunflower (*Helianthuus annuus* L.)	Oxalate oxidase activity	Increased resistance to *Sclerotinia sclerotiorum*	Hu et al. (2003)
	Potato (*Solanum tuberosum* L.)	Oxalate oxidase activity	Increased resistance to *Phytopthora infestans, Streptomyces reticuliscabiei.*	Schneider et al. (2002)
	Soybean (*Glycine max* L.)	Oxalate oxidase activity	Increased resistance to *Sclerotinia sclerotiorum*	Donaldson et al. (2001)
TaDhn-5	Arabidposis (*A. thaliana*)	Not determined	Increased tolerance to drought, salt	Brini et al. (2007)
TaWaox1a	Arabidposis (*A. thaliana*)	Mitochondrial alternative oxidase	Ameliorates ROS production under low temp	Sugie et al. (2006)
TaAs2	Arabidposis (*A. thaliana*)	Not determined	Regulation of plant development	Ma et al. (2007)
wft1, *wft2*	Perennial ryegrass (*Lolium perenne* L.)	Fructosyltransferase activity	Increased fructan content, freezing tolerance	Hisano et al. (2004)
PinA, PinB	Rice (*Oryza sativa* L.)	Grain softness	Reduced grain hardness	Krishnamurthy and Giroux (2001).
PinB	Apple (*Malus x domestica* Borkh)	Grain softness	Increased fungal resistance	Faize et al. (2004)
TaPP2Ac-1	Tobacco (*Nicotiana. tobacum* L.)	Ser/Thr and Tyr phopshatase	Increased tolerance to drought stress	Xu et al. (2007b)
WSOC1	Arabidposis (*A. thaliana*)	Flowering activator	Early flowering	Shitsukawa et al. (2007)
TaRAN1	Arabidposis (*A. thaliana*)	Protein transporter during mitosis	Regulation of mitosis in shoot apical and root meristem	Wang et al. (2006)

TaMADS1			Early flowering and altered floral organ development	Zhao et al. (2006)
WhUCP	Rice (*Oryza sativa* L.)	Mitochondrial uncoupling protein	Increase tolerance to cold treatment	Ozawa et al. (2006)
PMA80, PMA1959	Rice (*Oryza sativa* L.)	Water binding proteins	Increased drought and salt stress tolerance	Cheng et al. (2002)
Wcor15	Tobacco (*Nicotiana. tobacum* L.)	Not determined	Increased freezing tolerance	Shimamura et al. (2006)
TaTPC1	Yeast (*Saccharomyces cerevisieae*) Arabidposis (*A. thaliana*)	Ca^{2+} permeable channel	Restoration of Ca^{2+} channel deficient yeast mutant. Increased proportion of stomata with small aperture	Wang et al. (2005)
Wcs19	Arabidposis (*A. thaliana*)	Not determined	Increased freezing tolerance in cold-acclimated plants	Dong et al. (2002)
cys1	Tobacco (*Nicotiana. tobacum* L.)	Synthesis of Cys and glutathione	Enhanced tolerance to SO_2 exposure. Increase in Cu/Zn superoxide dismutase. Reduction in foliar and photooxidative damage	Youssefian et al. (2001)
wali5	Arabidposis (*A. thaliana*)	Not determined	Increased Al tolerance	Ezaki et al. (2000)
ALMT1	Barley (*Hordeun vulgare* L.)	Malate transporter protein	Increased Al tolerance in hydroponics and acid soil	Delhaize et al. (2004)
WCOR410	Strawberry (*Fragaria x anassassa*)	Not determined	Increased freezing tolerance in cold-acclimated plants	Houde et al. (2004)
TaPCS1	Arabidposis (*A. thaliana*)	Heavy metal detoxification	Regulation of long distance Cd^{2+} transport	Gong et al. (2003)
LMW-GS-MB1	Potato (*Solanum tuberosum* L.)	Viscoelastic properties	Hydrated flour with improved viscosity	Benmoussa et al. (2004)
TaHd1	Rice (*Oryza sativa* L.)	Not determined	Promoted heading under short day and delayed heading under long day	Nemoto et al. (2003)
WMn-SOD	Canola (*Brassica napus* L.)	Mitochondrial superoxide dismutase activity	Increased Al tolerance	Basu et al. (2001)

Similarly, overexpression of *pinA* genes for grain softness in transgenic hard wheat showed an increase in puroindoline content resulting in soft grain phenotype (Hogg et al. 2005; Martin et al. 2006) confirming the functional role of endogenous *pinA* controlling grain hardness. These studies clearly show functional attributes of over-expressing endogenous wheat genes and their contribution to grain quality and dough characteristics.

12.4.2 Transgenics and RNA Interference

In some instances, over expression may not necessarily provide a phenotype that determines functionality of an endogenous wheat gene. In addition, considerable resources are required, including cloning and complete of the target gene and inherent issues with regulatory constraints on using vector constructs in transformation (Shewry and Jones 2005; Bhalla 2006; Bhalla et al. 2006), limiting the capability for analysing a large number of genes simultaneously by over expression. Therefore, alternative transgenic approaches are required to elucidate gene function. RNA gene silencing provides a means whereby the phenotypic effect of suppressed transcription can be studied (Meins et al. 2005). RNA interference (RNAi) is a sequence-specific gene silencing mechanism whereby small interfering RNAs (siRNA) bind to target mRNA triggering the synthesis of double stranded RNA (dsRNA). The dsRNA are targets for cleaving by Dicer enzyme into subsequent siRNA molecules causing post-transcriptional gene silencing by mRNA degradation and methylation of homologous DNA (Meyer and Saedler 1996; Tijsterman et al. 2002; Denli and Hannon 2003; Baulcombe 2004; Meins et al. 2005). The application of RNAi for gene knockouts in wheat have confirmed gene function for *VRN1* (Loukoianov et al. 2005), *VRN2* (Yan et al. 2004), *GBSS* (Li et al. 2005), *SBE-IIa* (Regina et al. 2006), *SBE-IIb* (Regina et al. 2006), *EIN2* (Trevalla et al. 2006), *PDS* (Trevalla et al. 2006), *GPC* (Uauy et al. 2006) and *1Dx5* (Yue et al. 2008) and has been recently reviewed in Fu et al. (2007). Interestingly, overexpression and RNAi have complemented functional attributes of some wheat genes. For example, the overexpression of *1Dx5* in transgenic wheat produced highly cross-linked protein networks resulting in 'overstrong' dough quality characteristics (Darlington et al. 2003) whereas suppression of *1Dx5* by RNAi reduced gluten and diminished mixing quality (Yue et al. 2008). A major advantage of RNAi technology is the simultaneous silencing of all homoeologs or targeting siRNA to specific copies within the wheat genome and their individual or combined effects on plant phenotype.

12.4.3 Virus Induced Gene Silencing (VIGS)

Although the ability to develop gene knockouts provides a means whereby gene function can be determined, RNAi analysis requires the development of

transformed wheat plants. The limitations for determining the functional attributes of a large number of genes may be the efficiency of wheat transformation. This is particularly relevant in the post-genomic era where functional analysis of many wheat genes will assist in validating sequence annotation. Therefore, alternative strategies for more robust and efficient mechanisms for determining gene function must be considered and developed. VIGS is a transient system whereby a viral vector harbouring an endogenous sequence infects the host plant and triggers cytoplasmic degradation of RNA with homology to the target sequence. This strategy has been demonstrated to be effective in monocots using either Wheat Streak Mosaic Virus (WSMV) or Barley Stripe Mosaic Virus (BSMV) as a viral vector for post transcriptional gene silencing (Choi et al. 2000; Holzberg et al. 2002). VIGS has been effective in determining gene function for *Lr1* (Cloutier et al. 2007), *Lr21* (Scofield et al. 2005), phytoene desaturase (Holzberg et al. 2002; Scofield et al. 2005) and receptor-like kinases in stripe rust resistance (Zhou et al. 2007). Although VIGS offer a means for obtaining the rapid determination of gene function, the assays are generally transient and may be limited to silencing of genes to specific tissue where the virus infects.

12.4.4 Targeting Induced Local Lesions in Genomes

TILLING is a means whereby DNA sequence base differences can be efficiently detected amongst large population sizes. The first TILLING study in plants was done in Arabidopsis (McCallum et al. 2000) whereby a ethyl methanesulfonate (EMS) mutagenised population was screened and individuals selected for single base changes in chromomethylase genes. Details of TILLING methodology has been reviewed in Henikoff and Comai (2003), Henikoff et al. (2004), Gilchrist and Haughn (2005), Comai and Henikoff (2006). Briefly, DNA is extracted from individuals of the population, pooled and DNA of the target gene sequence is amplified. Following melting and reannealing of DNA, single base differences form heteroduplexes and detected using denaturing high-performance liquid chromatography (DHPLC) or cleaved using endonucleases such as CEL1 (Colbert et al. 2001) or ENDO1 (Triques et al. 2007) specific for single base mismatches followed by gel fragment analysis for mutation detection. Additionally, advances in mutation detection is now progressing beyond gel-based methodologies using non-fluorescence and fast capillary electrophoresis for higher throughput capabilities (Suzuki et al. 2008). TILLING approaches have recently been extended from model plants to crop species including barley (Caldwell et al. 2004), maize (Till et al. 2004) and rice (Till et al. 2007). Since these crops are diploid species, it was anticipated that TILLING in complex polyploid genomes such as hexaploid wheat would be more challenging because of redundant gene copies (Slade and Knauf 2005). Nevertheless, a TILLING study was done on hexaploid wheat where 94 mutations for the granule bound starch synthase (GBSS) genes was detected, identifying a near *waxy* phenotype with a triple homozygous mutation (Slade et al. 2005).

Therefore, the application of TILLING to large complex genome such as wheat provides alternative approaches to deciphering function of wheat genes.

The choice of systems to determine gene function relies on the resources available and the specific objectives of the study. VIGS would be ideally suited for validation of function for a large number of gene sequences in the post-genome era (such as ESTs) without the need to generate transgenic plants. However, VIGS is an unlikely strategy for developing commercial cultivars with modified gene expression. Alternatively, transgene overexpression, RNAi and TILLING offer the best approaches if both gene function and development of commercial wheat varieties are of simultaneous interest. TILLING would be the only option for developing new varieties for agricultural production zones in countries where the release of genetically modified wheat are constrained.

12.5 Application of Genomics to Wheat Breeding

Significant advances have been made in increasing our knowledge on phenotypic variation for cultivar development. Major wheat breeding programs world-wide are using DNA markers linked to genes controlling trait variation with some success during population enrichment. The knowledge of marker-trait associations for a range of agronomic disease and quality characteristics, marker availability and the strategies to deploy in a breeding program is summarized in recent reviews (Bonnett et al. 2005; Kuchel et al. 2005; William et al. 2007). However, low level DNA polymorphism between elite lines, recombination between marker and gene and the implementation of flanking markers separated by a large genetic distance can limit the effective use of tracking desirable alleles using linked DNA markers. Ideally, the gene responsible for trait variation would be of significant benefit in the application of marker-assisted selection for wheat improvement. The development of functionally associated markers (FAM) requires the isolation and functional characterization of genes known to control variation for a particular trait. The genomics tools and resources for genetic analysis, gene discovery and gene function provide the necessary requirements to identify FAM for wheat improvement. There are now at least 25 FAM reported for wheat (reviewed in Bagge et al. 2007) and represent diagnostic markers for allelic variation of important genes in major wheat breeding programs world-wide. The increasing development in wheat genomic resources such as those required for map or sequence-based cloning will provide a substantial increase in the number of FAM available to wheat breeding programs.

There is no doubt that breeding programs will need to effectively develop new varieties to mitigate the effects of global warming and climate change for future increase in wheat production. Adaptation to changing environments will be the key strategy to make significant advances and the genomics tools developed to date will play an important role. For example, FAM associated with vernalisation and photoperiod response will have a significant impact on assessing allelic variation in Australian genotypes to associate combinations with phenological

adaptation to maximise yield potential affected by changing environments (M. Francki, unpublished data). However, this will not be at the expense of using FAM or alternative markers as they are made available for numerous other traits associated with drought tolerance, extreme temperatures and other phenotypes (biotic and other abitoic stress) important for regional adaptation. The future application of genomics to identify new FAM for numerous traits will become more important to assist breeding high yielding wheat varieties adaptable to changing environments.

References

Allouis S, Moore G, Bellec A, Sharp R, Faivre P, Montimer K, Pateyron S, Foote T, Griffiths S, Caboche M, Chaloub B (2003) Construction and characterization of a hexaploid wheat (*Triticum aestivum* L.) BAC library from the reference germplasms 'Chinese Spring'. Cereal Res Commun 31:331–338

Almasy L, Blangero J (1998) Multipoint quantitative linkage analysis in general pedigrees. Am J Hum Genet 62:1198–1211

Altpeter F, Popelka JC, Wieser H (2004) Stable expression of 1Dx5 and 1Dy10 high-molecular weight glutenin subunit genes in transgenic rye drastically increases the polymeric glutelin fraction in rye flour. Plant Mol Biol 54:783–792

Alvarez ML, Guelman S, Halford NG, Lustig S, Reggiardo MI, Ryabushkina N, Shewry P, Stein J, Vallejos RH (2000) Silencing of HMW glutenins in transgenic wheat expressing extra HMW subunits. Theor Appl Genet 100:319–327

Appels R, Francki M, Chibbar R (2003) Advances in cereal functional genomics. Funct Integr Genomics 3:1–24

Bagge M, Xia XC, Lubberstedt T (2007) Functional markers in wheat- commentary. Curr Opin Plant Biol 10:211–216

Barro F, Rooke L, Bekes F, Gras P, Tatham AS, Fido R, Lazzeri PA, Shewry PR, Barcelo P (1997) Transformation of wheat with high molecular weight subunit genes results in improved functional properties. Nat Biotech 15:1295–1299

Basu U, Good AG, Taylor GJ (2001) Transgenic *Brassica napus* plants overexpressing aluminium-induced mitochondrial manganese superoxide dismutase cDNA are resistant to aluminium. Plant Cell Environ 24:1269–1278

Baulcombe D (2004) RNA silencing in plants. Nature 431:356–363

Benmoussa M, Vezina LP, Page M, Gelinas P, Yelle S, Laberge S (2004) Potato flour viscosity improvement is associated with the expression of a wheat LMW-glutenin gene. Biotech Bioeng 87:495–500

Bhalla PL (2006) Genetic engineering of wheat- current challenges and opportunities. Trends Biotech 24:305–311

Bhalla PL, Ottenhof HH, Singh MB (2006) Wheat transformation- an update of recent progress. Euphytica 149:353–366

Blechl A, Lin J, Nguyen S, Chan R, Anderson OD, Dupont FM (2007) Transgenic wheats with elevated levels of Dx5 and/or Dy10 high molecular weight glutenin subunits yield doughs with increased mixing strength and tolerance. J Cereal Sci 45:172–183

Bonnett DG, Rebetzke GJ, Spielmeyer W (2005) Strategies for efficient implementation of molecular markers in wheta breeding. Mol Breed 15:75–85

Bossolini E, Wicker T, Knobel PA, Keller B (2007) Comparison of orthologous loci from small grass genomes Brachypodium and rice: implications for wheat genomics and grass genome annotation. Plant J 49:704–717

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177

Brini F, Hanin M, Lumbreras V, Amara I, Khoudi H, Hassairi A, Pages M, Masmoudi K (2007) Overexpression of wheat dehydrin DHN-5 enhances tolerance to salt and osmotic stress in *Arabidopsis thaliana*. Plant Cell Rep 26:2017–2026

Caldwell D, McCallum N, Shaw P, Muehlbauer GJ, Marshall DF, Waugh R (2004) A structured mutant population for forward and reverse genetics in barley (*Hordeum vulgare* L.). Plant J 40:143–150

Caldwell KS, Russell J, Langridge P, Powell W (2006) Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. Genetics 172:557–567

Cenci A, Chantret N, Kong X, Gu Y, Anderson OD, Fahima T, Distelfeld A, Dubcovsky J (2003) Construction and characterization of a half million clone BAC library of durum wheat (*Triticum turgidum* ssp. *durum*). Theor Appl Genet 107:931–939

Chalmers KJ, Campbell AW, Kretschmer J, Karakousis A, Henschke PH, Pierens S, Harker N, Pallotta M, Cornish GB, Shariflou MR, Rampling LR, McLauchlan A, Daggard G, Sharp PJ, Holton TA, Sutherland MW, Appels R, Langridge P (2001) Construction of three linkage maps in bread wheat, *Triticum aestivum*. Aust J Agric Res 52:1089–1119

Chao S, Sharp PJ, Worland AJ, Warham EJ, Koebner RMD, Gale MD (1989) RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. Theor Appl Genet 78:495–504

Chao S, Lazo GR, You F, Crossman CC, Hummel DD, Lui N, Laudencia-Chingcuanco D, Anderson JA, Close TJ, Dubcovsky J, Gill BS, Gill KS, Gustafson JP, Kianian SF, Lapitan NLV, Nguyen HT, Sorrells ME, McGuire PE, Qulaset CO, Anderson OD (2006) Use of large-scale Triticeae expressed sequence tag resource to reveal gene expression profiles in hexaploid wheat (*Triticum aestivum* L.). Genome 49:531–544

Cheng ZQ, Targolli J, Huang XQ, Wu R (2002) Wheat LEA genes *PMA80* and *PMA1959*, enhance dehydration tolerance of transgenic rice (Oryza sativa L.). Mol Breed 10:71–82

Choi I-R, Stenger DC, Morris TJ, French R (2000) A plant virus vector for systemic expression of foreign genes in cereals. Plant J 23:547–555

Cloutier S, McCallum BD, Loutre C, Banks TW, Wicker T, Feiullet C, Keller B, Jordan MC (2007) Leaf rust resistance gene *Lr1*, isolated from bread wheat (*Triticum aestivum* L.) is a member of the large psr567 gene family. Plant Mol Biol 65:93–106

Colbert T, Till BJ, Tompa R, Reynolds S, Steine MN, Yeung AT, McCallum CM, Comai L, Henikoff S (2001) Highthroughput screening for induced point mutations. Plant Physiol 126:480–484

Comai L, Henikoff S (2006) TILLING: practical single-nucleotide mutation discovery. Plant J 45:684–694

Crepieux S, Lebreton C, Servin B, Charmet G (2004) Quantitative trait loci (QTL) detection in multicross inbred designs: recovering QTL identical-by-descent status information by marker data. Genetics 168:1737–1749

Darlington H, Fido R, Tatham AS, Jones H, Salmon SE, Shewry PR (2003) Milling and baking properties of field grown wheat expressing HMW subunit transgenes. J Cereal Sci 38:301–306

Delhaize E, Ryan PR, Hebb DM, Yamamoto Y, Sasaki T, Matsumoto H (2004) Engineering high-level aluminium tolerance in barley with the *ALMT1* gene. Proc Natl Acad Sci 101:15249–15254

Denli AM, Hannon GJ (2003) RNAi: an ever-growing puzzle. Trends Biochem Sci 28:196–201

Devos KM, Gale MD (1997) Comparative genetics in the grasses. Plant Mol Biol 35:3–15

Devos KM, Atkinson MD, Chinoy CN, Liu C, Gale MD (1992) RFLP based genetic map of the homoeologous group 3 chromosomes of wheat and rye. Theor Appl Genet 83:931–939

Devos KM, Millan T, Gale MD (1993) Comparative RFLP maps of homoeologous group 2 chromosomes of wheat, rye, and barley. Theor Appl Genet 85:784–792

Devos KM, Dubcovsky J, Dvorak J, Chinoy CN, Gale MD (1995) Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination. Theor Appl Genet 91:282–288

Devos KM, Ma J, Pontaroli AC, Pratt LH, Bennetzen JL (2005) Analysis and mapping of randomly chosen bacterial artificial chromosome clones from hexaploid bread wheat. Proc Natl Acad Sci 102:19243–19248

Donaldson PA, Anderson T, Lane BG, Davidson AL, Simmonds DH (2001) Soybean plants expressing an active oligomeric oxalate oxidase from the wheat *gf-2.8* (germin) gene are resistant to the oxalate secreting pathogen *Sclerotinia sclerotiorum*. Physiol Mol Plant Pathol 59:297–307

Dong CN, Danyluk J, Wilson KE, Pocock T, Huner NPA, Sarhan F (2002) Cold-regulated cereal chloroplast late embryogenesis abundant-like proteins. Molecular characterization and functional analysis. Plant Physiol 129:1368–1381

Draper J, Mur LAJ, Jenkins G, Ghosh-Biswas GC, Bablak P, Hasterok R, Routledge APM (2001) *Brachypodium distachyon*. A new model system for functional genomics in grasses. Plant Physiol 127:1539–1555

Erayman M, Sandhu D, Sidhu D, Dilbirligi M, Baenziger PS, Gill KS (2004) Demarcating the gene-rich regions of the wheat genome. Nucleic Acids Res 32:3546–3565

Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W (2002) Isolation of EST-derived microsatellite markers for genotpying the A and B genomes of wheat. Theor Appl Genet 104:399–407

Ezaki B, Gardner RC, Ezaki Y, Matsumoto H (2000) Expression of aluminium-induced genes in transgenic Arabidopsis plants can ameliorate aluminium stress and/or oxidative stress. Plant Physiol 122:657–665

Faize M, Sourice S, Dupuis F, Parisi L, Gautier MF, Chevreau E (2004) Expression of wheat puroindoline-b reduces scab resistance in transgenic apple (Malus × Domestica Borkh). Plant Sci 167:347–354

Feuillet C, Travella S, Stein N, Albar L, Nublat A, Keller B (2003) Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome. Proc Natl Acad Sci. 100:15253–15258

Foote T, Griffiths S, Allouis S, Moore G (2004) Construction and analysis of a BAC library in the grass *Brachypodium sylvaticum* and its use as a tool to bridge the gap between rice and wheat in elucidating gene content. Funct Integr Genomics 4:26–33

Francki M, Appels R (2007) Comparative genomics and crop improvement. In: James R. Brown (ed) Comparative genomics: fundamental and applied perspectives. CRC/Taylor & Francis, New York, USA, pp 323–342

Francki MG, Carter M, Ryan K, Hunter A, Bellgard M, Appels R (2004) Comparative organization of wheat homoeologous group 3S and 7L using wheat-rice synteny and identification of potential markers for genes controlling xanthophyll content in wheat. Funct Integr Genomics 4:118–130

Francki MG, Walker E, Forster JW, Spangenburg G, Appels R (2006) Fructosyltransferase and invertase genes evolved by gene duplication and rearrangement: rice, perennial ryegrass and wheat gene families. Genome 49:1081–1091

Fu D, Uauy C, Blechl A, Dubcovsky J (2007) RNA interference for wheat functional gene analysis. Transgenic Res 16:689–701

Gadaleta A, Mangini, Mule G, Blanco A (2007) Charaterization of dicnulceotide and trinucleotide EST-SSR derived microsatellites in the wheat genome. Euphytica 153:73–85

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci USA 95:1971–1974

Gao LF, Jing RL, Huo NX, Li Y, Zhou RH, Chang XP, Tang JF, Ma ZY, Jia JZ (2004) One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. Theor Appl Genet 108:1392–1400

Gilchrist EJ, Haughn GW (2005) TILLING without a plough: a new method with applications for reverse genetics. Curr Opin Plant Biol 8:211–215

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L ssp japonica). Science 296:92–100

Gong JM, Lee DA, Schroeder JI (2003) Long distance root-to-shoot transport of phytochelatins and cadmium in Arabidopsis. Proc Natl Acad Sci 100:10118–10123

Gupta P, Balyan H, Edwards K, Isaac P, Korzun V, Röder M, Gautier M-F, Joudrier P, Schlatter A, Dubcovsky J, De la Pena R, Khairallah M, Penner G, Hayden M, Sharp P, Keller B, Wang R, Hardouin J, Jack P, Leroy P (2002) Genetic mapping of 66 new microsatellite (SSR) loci in bread wheat. Theor Appl Genet 105:413–422

Gupta PK, Rustgi S, Sharma S, Singh R, Kumay N, Blayan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Gen Genomics 270:315–323
Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: Present status and future prospects. Plant Mol Biol 57:461–485
Handa H, Namiki N, Xu D, Ban T (2008) Dissection of the FHB resistance QTL in the short arm of wheat chromosome 2D using comparative genomic approach: from QTL to candidate gene. Mol Breed. doi: 10.1007/s11032-008-9157-7
Hayden MJ, Nguyen TM, Waterman A, McMichael GL, Chalmers KJ (2008a) Application of multiplex-ready PCR for fluoresence-based SSR genotyping in wheat and barley. Mol Breed 21:271–281
Hayden MJ, Nguyen TM, Waterman A, Chalmers KJ (2008b) Multiplex-ready PCR: a new method for multiplexed SSR and SNP genotyping. BMC Genomics 9:80
He GY, Jones HD, D'Ovidio R, Masci S, Chen MJ, West J, Butow B, Anderson OD, Lazzeri P, Fido R, Shewry PR (2005) Expression of an extended HMW subunit in transgenic wheat and the effect on dough mixing properties. J Cereal Sci 42:225–231
Henikoff S, Comai L (2003) Single-nucleotide mutations for plant functional genomics. Ann Rev Plant Biol 54:375–401
Henikoff S, Till BJ, Comai L (2004) TILLING. Traditional mutagenesis meets functional genomics. Plant Physiol 135:630–636
Hisano H, Kanazawa A, Kawakami A, Yoshida M, Shimamoto Y, Yamada T (2004) Transgenic perennial ryegrass plants expressing wheat fructosyltransferase genes accumulate increased amounts of fructan and acquire increased tolerance on a cellular level to freezing. Plant Sci 167:861–868
Hogg AC, Beecher B, Martin JM, Meyer F, Talbert L, Lanning S, Giroux MJ (2005) Hard wheat milling and bread baking traits affected by the seed-specific overexpression of puroindolines. Crop Sci 45:871–878
Holzberg S, Brosio P, Gross C, Pogue GP (2002) Barley stripe mosaic virus-induced gene silencing in a monocot plant. Plant J 30:315–327
Houde M, Dallaire S, N'Dong D, Sarhan F (2004) Overexpression of the acidic dehydrin WCOR410 improves freezing tolerance in transgenic strawberry leaves. Plant Biotech J 2:381–387
Hu X, Bidney DL, Yalpani N, Duvick JP, Crasta O, Folkerts O, Lu GH (2003) Overexpression of a gene encoding hydrogen peroxide generating oxalate oxidase evokes defense responses in sunflower. Plant Physiol 133:170–181
Huang L, Brooks SA, Li W, Feller JP, Trick HN, Gill BS (2003) Map-based cloning of leaf rust resistance gene *Lr21* from the large and polyploid genome of bread wheat. Genetics 164:655–664
International Rice Genome Sequencing Project (2005) The map based sequence of the rice genome. Nature 436:793–800
Janda J, Bartos J, Safar J, Kubalakova M, Valarik M, Cihakikova J, Simkova H, Caboche M, Sourdille P, Bernard M, Chalhoub B, Dolezel J (2004) Construction of a subgenomic BAC library specific for chromosome 1D, 4D and 6D of hexaploid wheat. Theor Appl Genet 109:1337–1345
Janda J, Safar J, Kubalakova M, Bartos J, Kovarova P, Suchankova P, Pateyron S, Cihalikova J, Sourdille P, Simkova H, Faivre-Rampant P, Hribova E, Bernard M, Lukaszewski A, Dolezel J, Chalhoub B (2006) Advanced resources for plant genomics: a BAC library specific for the short arm of wheat chromosome 1B. Plant J 47:977–986
Jia J, Devos KM, Chao S, Miller TE, Reader SM, Gale MD (1996) RFLP based maps of the homoeologous group 6 chromosomes of wheat and their application in the tagging of *Pm12*, a powdery mildew resistance gene transferred from *Aegilops speltoides* to wheat. Theor Appl Genet 92:559–565
Kantety RV, La Rota M, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 48:501–510

Kellog EA (2001) Evolutionary history of the grasses. Plant Physiol 125:1198–1205
Krishnamurthy K, Giroux MJ (2001) Expression of wheat purindoline genes in transgenic rice enhances grain softness. Nature Biotech 19:162–166
Kuchel H, Ye G, Fox R, Jefferies S (2005) Genetic and economic analysis of a targeted marker assisted wheat breeding strategy. Mol Breed 16:67–78
La Rota M, Sorrells ME (2004) Comparative DNA sequence analysis of mapped wheat ESTs reveals the complexity of genome relationships between rice and wheat. Funct Integr Gen 4:34–46
Lantican MA, Pingali PL, Rajaram S (2002) Are marginal wheat environments catching up?. In: Ekboir J (ed) CIMMYT 2000–2001 world wheat overview and outlook: developing no-till packages for small-scale farmers. CIMMYT, Mexico, DF
Lazo GR, Chao S, Hummel DD, Edwards H, Crossman CC et al. (2004) Development of an expressed sequence tag (EST) resource for wheat (*Triticum aestivum* L): EST generation, unigene analysis, probe selection and bioinformatics for a 16,000-locus bin-delineated map. Genetics 168:585–593
Li C, Ni P, Francki M, Hunter A, Zhang Y, Schibeci D, Li H, Tarr A, Wang J, Cakir M, Yu J, Bellgard M, Lance R, Appels R (2004) Genes controlling seed dormancy and pre-harvest sprouting in a rice-wheat barley comparison. Funct Integr Genomics 4:84–93
Li JR, Zhao W, Li QZ, Ye XG, An BY, Li X, Zhang XS (2005) RNA silencing of *Waxy* gene results in low levels of amylose in the seeds of transgenic wheat (*Triticum aestivum* L.). Acta Genet Sin 32:846–854
Liang HY, Maynard CA, Allen RD, Powell WA (2001) Increasing *Septoria musiva* resistance in transgenic hybrid poplar leaves expressing a wheat oxalate oxidase gene. Plant Mol Biol 45:619–629
Lijavetzky D, Muzzi G, Wicker T, Keller B, Wing R, Dubcovsky J (1999) Construction and characterization of a bacterial artificial chromosome (BAC) library for the A genome of wheat. Genome 42:1176–1182
Ling P, Chen XM (2005) Construction of a hexaploid wheat (*Triticum aestivum* L.) bacterial artificial chromosome library for cloning genes for stripe rust resistance. Genome 48:1028–1036
Liu S, Anderson JA (2003) Targeted molecular mapping of a major wheat QTL for Fusarium head blight resistance using wheat ESTs and synteny with rice. Genome 46:817–823
Liu YG, Nagaki K, Fujita M, Kawaura K, Uozumi M, Ogihara Y (2000) Development of an efficient maintenance and screening system for large-insert genomic DNA libraries of hexaploid wheat in a transformation –competent artificial chromosome (TAC) vector. Plant J 23:687–695
Loukoianov A, Yan L, Blechl A, Sanchez A, Dubcovsky J (2005) Regulation of *VRN-1* vernalization genes in normal and transgenic polyploid wheat. Plant Physiol 138:2364–2373
Luo MC, Thomas C, You FM, Hsiao J, Ouyang S, Buell CR, Malandro M, McGuire PE, Anderson OD, Dvorak J (2003) High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics 82:378–389
Ma XJ, Tao Y, Zhao XY, Zhang XS (2007) Wheat *TaAS2*, a member of LOB family, affects the abaxial–abaxial polarity of leaves in transgenic Arabidopsis. Plant Sci 172:181–188
MacKay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. Trends Plant Sci 12:57–63
Martin JM, Meyer FD, Smidansky ED, Wanjugi H, Blechl AE, Giroux MJ (2006) Complementation of the *pina* (null) allele with the wild type *Pina* sequence restores a soft phenotype in transgenic wheat. Theor Appl Genet 113:1563–1570
Masci S, D'Ovidio R, Scossa F, Patacchini C, Lafiandra D, Anderson OD, Blechl AE (2003) Production and characterization of a transgenic bread wheat line over-expressing a low molecular weight glutenin subunit gene. Mol Breed 12:209–222
McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeted screening for induced mutations. Nat Biotech 18:455–457
McFadden ES, Sears ER (1946) The origin of *Triticum spelta* and its free threshing hexaploid relatives. J Hered 37:81–89

Meins F, S-Ammour A, Blevins T (2005). RNA silencing systems and their relevance in plant development. Ann Rev Cell Dev Biol 21:297–318

Meyer P, Saedler H (1996) Homology-dependent gene silencing in plants. Ann Rev Plant Physiol Mol Biol 47:23–48

Miftahudin, Ross K, Ma X-F, Mahmoud AA, Layton J, Rodriguez-Milla MA, Chikmawati T, Ramalingam J, Feril O, Pathan MS, Surlan Momirovic G, Kim S, Chema K, Fang P, Haule L, Struxness H, Birkes J, Yaghoubian C, Skinner R, McAllister J, Nguyen V, Qi LL, Echalier B, Gill BS, Linkiewicz AM, Dubcovsky J, Akhunov ED, Dvorák J, Dilbirligi M, Gill KS, Peng JH, Lapitan NLV, Bermudez-Kandianis CE, Sorrells ME, Hossain KG, Kalavacharla V, Kianian SF, Lazo GR, Chao S, Anderson OD, Gonzalez-Hernandez J, Conley EJ, Anderson JA, Choi D-W, Fenton RD, Close TJ, McGuire PE, Qualset CO, Nguyen HT, Gustafson JP (2004) Analysis of expressed sequence tag loci on wheat chromosome group 4. Genetics 168:651–663

Mochida K, Yamazaki Y, Ogihara Y (2004) Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. Mol Gen Genomics 270:371–377

Morimoto R, Kosugi T, Nakamura C, Takumi S (2005) Intragenic diversity and functional conservation of the three homoeologous loci of the KN1-type homeobox gene *Wknox1* in common wheat. Plant Mol Biol 57:907–924

Moullet O, Zhang HB, Lagudah ES (1999) Construction and characterisation of a large DNA insert library from the D genome of wheat. Theor Appl Genet 99:305–313

Mullan DJ, Colmer TM, Francki MG (2007) *Arabidopsis*-rice-wheat orthologues for Na^+ transport and transcript analysis in wheat-*L. elongatum* aneuploids under salt stress. Mol Genet Genomics 277:199–212

Munkvold JD, Greene RA, Bermudez-Kandianis CE, La Rota CM, Edwards H, Sorrells SF, Dake T, Benscher D, Kantety R, Linkiewicz AM, Dubcovsky J, Akhunov ED, Dvorák J, Miftahudin, Gustafson JP, Pathan MS, Nguyen HT, Matthews DE, Chao S, Lazo GR, Hummel DD, Anderson OD, Anderson JA, Gonzalez-Hernandez JL, Peng JH, Lapitan N, Qi LL, Echalier B, Gill BS, Hossain KG, Kalavacharla V, Kianian SF, Sandhu D, Erayman M, Gill KS, McGuire PE, Qualset CO, Sorrells ME (2004) Group 3 chromosome bin maps of wheat and their relationship to rice chromosome 1. Genetics 168:639–650

Nemoto Y, Kisaka M, Fuse T, Yano M, Ogihara Y (2003) Characterization and functional analysis of three wheat genes with homology to the CONSTANS flowering time gene in transgenic rice. Plant J 36:82–93

Nilmalgoda SD, Cloutier S, Walichnowski AZ (2003) Construction and characterization of a bacterial artificial chromosome (BAC) library of hexaploid wheat (*Triticum aestivum* L.) and validation of genome coverage using locus-specific primers. Genome 46:870–878

Ozawa K, Murayama S, Kobayashi-Uehara A, Handa H (2006) Overexpression of wheat mitochondrial uncoupling protein in rice plants confers tolerances to oxidative stresses promoted by exogenous hydrogen peroxidase and low temperature. Mol Breed 18:51–56

Pestsova E, Ganal MW, Roder MS (2000) Isolation and mapping of microsatellite markers specific for the D genome of bread wheat. Genome 43:689–697

Qi LL, Echalier B, Chao S, Lazo GR, Butler GE, Anderson OD, Akhunov ED, Dvorák J, Linkiewicz AM, Ratnasiri A, Dubcovsky J, Bermudez-Kandianis CE, Greene RA, Kantety R, La Rota CM, Munkvold JD, Sorrells SF, Sorrells ME, Dilbirligi M, Sidhu D, Erayman M, Randhawa HS, Sandhu D, Bondareva SN, Gill KS, Mahmoud AA, Ma X-F, Miftahudin, Gustafson JP, Conley EJ, Nduati V, Gonzalez-Hernandez JL, Anderson JA, Peng JH, Lapitan NLV, Hossain KG, Kalavacharla V, Kianian SF, Pathan MS, Zhang DS, Nguyen HT, Choi D-W, Fenton RD, Close TJ, McGuire PE, Qualset CO, Gill BS (2004) A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. Genetics 168:701–712

Qi LL, Echalier B, Friebe B, Gill BS (2003) Molecular characterization of a set of wheat deletion stocks for use in chromosome bin mapping of ESTs. Funct Integr Genomics 3:39–55

Rafalski A, Morgante M (2004) Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. Trends Genet 20:103–111

Ramputh AI, Arnason JT, Cass L, Simmonds JA (2002) Reduced herbivory of the European corn borer (*Ostrinia nubilalis*) on corn transformed with germin, a wheat oxalate oxidase gene. Plant Sci 162:431–440

Randhawa HS, Dilbirligi M, Sidhu D, Erayman M, Sandhu D, Bondareva S, Chao S, Lazo GR, Anderson OD, Miftahudin, Gustafson JP, Echalier B, Qi LL, Gill BS, Akhunov ED, Dvorák J, Linkiewicz AM, Ratnasiri A, Dubcovsky J, Bermudez-Kandianis CE, Greene RA, Sorrells ME, Conley EJ, Anderson JA, Peng JH, Lapitan NLV, Hossain KG, Kalavacharla V, Kianian SF, Pathan MS, Nguyen HT, Endo TR, Close TJ, McGuire PE, Qualset CO, Gill KS (2004) Deletion mapping of homoeologous group 6-specific wheat expressed sequence tags. Genetics 168:677–686

Ravel C, Praud S, Murigneux A, Canaguier A, Sapet F, Samson D, Balfourier F, Dufour P, Chalhoub B, Brunel D, Linossier L, Beckert M, Charmet G (2006) Single-nucleotide polymorphism frequency in a set of selected lines of bread wheat (*Triticum aestivum* L.). Genome 49:1131–1139

Ravel C, Praud S, Canaguier A, Dufour P, Giancola S, Balfourier F, Chalhoub B, Brunel D, Linossier L, Dardevet M, Beckert M, Rousset M, Murigneux A, Charmet G (2007) DNA sequence polymorphisms and their application to bread wheat quality. Euphytica 158:331–336

Regina A, Bird A, Topping D, Bowden S, Freeman J, Barsby T, Kosar-Hashemi B, Li Z, Rahman S, Morrell M (2006) High-amylose wheat generated by RNA interference improves indices of large-bowel health in rats. Proc Natl Acad Sci USA 103:3546–3551

Remington DL, Purugganan MD (2003) Candidate genes, quantitative trait loci and functional trait evolution in plants. Int J Plant Sci 164:S7–S20

Roder MS, Korzun V, Wendehake K, Plaschke J, Tixier M-H, Leroy P, Ganal MW (1998) A microsatellite map of wheat. Genetics 149:2007–2023

Rooke L, Barro F, Tatham AS, Fido R, Steele S, Bekes F, Gras P, Martin A, Lazzeri PA, Shewry PR, Barcelo P (1999) Altered functional properties of tritordeum by transformation with HMW glutenin subunit genes. Theor Appl Genet 99:851–858

Safar J, Bartos J, Janda J, Bellec A, Kubalakova M, Valarik M, Pateyron S, Weiserova J, Tuskova R, Cihalikova J, Vrana J, Simkova H, Faivre-Rampant P, Sourdille P, Caboche M, Bernard M, Dolezel J, Chalhoub B (2004) Dissecting large and complex genomes: flow sorting and BAC cloning of individual chromosomes from bread wheat. Plant J. 39:960–968

Sangtong V, Moran DL, Chikwamba R, Wang K, Woodman-Clikeman W, Long MJ, Lee M, Scott MP (2002) Expression and inheritance of the wheat Glu-1DX5 gene in transgenic maize. Theor Appl Genet 105:937–945

Schneider M, Droz E, Malnoe P, Chatot C, Bonnel E, Metraux JP (2002) Transgenic potato plants expressing oxalate oxidase have increased resistance to oomycete and bacterial pathogens. Potato Res 45:177–185

Scofield SR, Huang L, Brandt AS, Gill BS (2005) Development of a virus-induced gene-silencing system for hexaploid wheat and its use in functional analysis of the *Lr21*-mediated leaf rust resistance pathway. Plant Physiol 138:2165–2173

Scott MP, Peterson JM, Moran DL, Sangtong V, Smith L (2007) A wheat genomic DNA fragment reduces pollen transmission of maize transgenes by reducing pollen viability. Trans Res 16:629–643

Shankar M, Walker E, Golzar H, Loughman R, Wilson RE, Francki MG (2008) QTL for seedling and adult plant resistance to *Stagonospora nodorum* in wheat. Phytopathol 98:886–893

Shen X, Francki MG, Ohm HW (2006) A resistance-like gene identified by EST mapping and its association with a QTL controlling *Fusarium* head blight infection on wheat chromosome 3BS. Genome 49:631–635

Shewry PR, Jones HD (2005) Transgenic wheat; where do we stand after the first 12 years? Ann Appl Biol 147:1–14

Shimamura C, Ohno R, Nakamura C, Takumi S (2006) Improvement of freezing tolerance in tobacco plants expressing a cold-responsive and chloroplast-targeting protein *WCOR15* of wheat. J Plant Physiol 163:213–219

Shitsukawa N, Ikari C, Mitsuya T, Sakiyama T, Ishikawa A, Takumi S, Murai K (2007) Wheat *SOC1* functions independently of WAP1/VRN1, an integrator of vernalization and photoperiod flowering promotion pathways. Phyisiol Plant 130:627–636

Singh NK, Dalal V, Batra K, Singh BK, Chitra G, Singh A, Ghazi IA, Yadav M, Pandit A, Dixit R, Singh PK, Singh H, Koundal KR, Gaikwas K, Mohapatra T, Sharma TR (2007) Single-copy genes define a conserved order between rice and wheat for understanding differences caused by duplication, deletion, and transposition of genes. Funct Integr Genomics 7:17–35

Slade AJ, Knauf VC (2005) TILLING moves beyond functional genomics into crop improvement. Transgenic Res 14:109–115

Slade AJ, Fuerstenberg SI, Loeffler D, Steine MN, Facciotti D (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nat Biotech 23:75–81

Somers DJ, Kirikpatrick R, Moniwa M, Walsh A (2003) Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. Genome 46:431–437

Somers DJ, Isaac P, Edwards K (2004) A high-density mircrosatellite map for bread wheat (*Triticum aestivum* L.) Genome 109:1105–1114

Somers DJ, Banks T, DePauw R, Fox S, Clarke J, Pozniak C, McCartney C (2007) Genome-wide linkage disequilibrium analysis in bread wheat and durum wheat. Genome 50:557–567

Soreng RJ, Davis JI (1998) Phylogenetics and character evolution in the grass family. The Bot Rev 64:1–47

Sorrells ME, La Rota M, Bermudez-Kandianis CE, Greene RA, Kantety R, Munkvold JD, Miftahudin, Mahmoud A, Ma X, Gustafson JP, Qi LL, Echalier B, Gill BS, Matthews DE, Lazo GR, Chao S, Anderson OD, Edwards H, Linkiewicz AM, Dubcovsky J, Akhunov ED, Dvorak J, Zhang D, Nguyen HT, Peng J, Lapitan NLV, Gonzalez-Hernandez JL, Anderson JA, Hossain K, Kalavacharla V, Kianian SF, Choi D-W, Close TJ, Dilbirligi M, Gill KS, Steber C, Walker-Simmons MK, McGuire PE, Qualset CO (2003) Comparative DNA sequence analysis of wheat and rice genomes. Genome Res 13:1818–1827

Stein N (2007) Triticeae genomics: advances in sequence analysis of large genome cereal crops. Chrom Res 15:21–31

Sugie A, Naydenov N, Mizuno N, Nakamura C, Takumi S (2006) Overexpression of wheat alternative oxidase *Waox1a* alters respiration capacity and response to reactive oxygen species under low temperature in transgenic Arabidopsis. Genes Genet Syst 81:349–354

Suzuki T, Eiguchi M, Kumamaru T, Satoh H, Matsusaka H, Moriguchi K, Nagato Y, Kurata N (2008) MNU-induced mutant pools and high performance TILLING enable finding of any gene mutation in rice. Mol Gen Genet 279:213–223

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 498:796–815

Tijsterman M, Ketting RF, Plasterk RHA (2002) The genetics of RNA silencing. Ann Rev Genet 36:489–519

Till BJ, Reynolds SH, Weil C, Springer N, Burtner C, Young K, Bowers E, Codomo CA, Enns LC, Odden AR, Greene EA, Comai L, Henikoff S (2004) Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biol 4:12

Till BJ, Cooper J, Tai TH, Colowit P, Greene EA, Henikoff S, Comai L (2007) Discovery of chemically induced mutations in rice by TILLING. BMC Plant Biol 7:19

Tosi P, D'Ovidio R, Napier JA, Bekes F, Shewry PR (2004) Expression of epitope-tagged LMW glutenin subunits in the starchy endosperm of transgenic wheat and their incorporation into the glutenin polymers. Theor Appl Genet 108:468–476

Tosi P, Masci S, Giovangrossi A, D'Ovidio R, Bekes F, Larroque O, Napier J, Shewry P (2005) Modification of the low molecular weight (LMW) glutenin composition of transgenic durum wheat: Effects on glutenin polymer size and gluten functionality. Mol Breed 16:113–126

Trevalla S, Klimm TE, Keller B (2006) RNA interference-based gene silencing as an efficient tool for functional genomics in hexaploid bread wheat. Plant Physiol142:6–20

Triques K, Sturbois B, Gallais S, Dalmais M, Chauvin S, Clepet C, Aubourg S, Rameau C, Caboche M, Bendahmane A (2007) Characterization of *Arabidopsis thaliana* mismatch specific endonucleases: application to mutation discovery by TILLING in pea. Plant J 51:1116–1125

Uauy C, Distalfeld A, Fahima T, Blechl A, Dubcovsky J (2006) A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. Science 314:1298–1301

Uphaus J, Walker E, Shankar M, Golzar H, Loughman R, Francki M, Ohm H (2007) Quantitative trait loci identified for resistance to Stagonospora glume blotch in wheat in the USA and Australia. Crop Sci 47:1813–1822

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995). AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Wang JR, Wei YM, Yan ZH, Zheng YL (2007) Sequence variations and haplotype identification of wheat dimeric alpha-amylase inhibitor genes in einkorn wheats. Biochem Genet 45:803–814

Wang X, Xu YY, Han Y, Bao SL, Du JZ, Yuan M, Xu ZH, Chong K (2006) Overexpression of RAN1 in rice and Arabidopsis alters primordial meristem, mitotic progress and sensitivity to auxin. Plant Physiol 140:91–101

Wang YJ, Yu JN, Chen T, Zhang ZG, Hao YJ, Zhang JS, Chen SY (2005) Functional analysis of a putative Ca^{2+} channel gene *TaTPC1* from wheat. J Exp Bot 56:3051–3060

William HM, Trethowan R, Crosby-Galvan EM (2007) Wheat breeding assisted by markers: CIMMYT's experience. Euphytica 157:307–319

Wolfe KH, Gouy M, Yang YW, Sharp PM, Li WH (1989) Date of the monocot–dicot divergence estimated from chloroplast DNA sequence data. Proc Natl Acad Sci U S A 86:6201–6205

Xie DX, Devos KM, Moore G, Gale MD (1993) RFLP based genetic maps of the homoeologous group 5 chromosomes of bread wheat (*Triticum aestivum* L.) Theor Appl Genet 87:70–74

Xu CY, Jing RL, Mao XG, Jia XY, Chang XP (2007b) A wheat (*Triticum aestivum*) protein phosphatase 2A catalytic subunit gene provides enhanced drought tolerance in tobacco. Ann Bot 99:439–450

Xu ZS, Xia LQ, Chen M, Cheng XG, Zhang RY, Li LC, Zhao YX, Lu Y, Ni ZY, Liu L, Qiu ZG, Ma YZ (2007a) Isolation and molecular characterization of the *Triticum aestivum* L. ethylene-responsive factor 1 (*TaERF1*) that increases multiple stress tolerance. Plant Mol Biol 65:719–732

Yahiaoui N, Srichumpa P, Dudler R, Keller B (2004) Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene *Pm3b* from hexaploid wheat. Plant J 37:528–538

Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J (2003) Positional cloning of the wheat vernalization gene *VRN1*. Proc Natl Acad Sci U S A. 100:6263–6268

Yan L, Loukoianov A, Blechl A, Tranquili G, Ramakrishna W, San Miguel P, Bennetzen JL, Echenique V, Dubcovsky J (2004) The wheat gene *VRN2* gene is a flowering repressor down-regulated by vernalization. Science 303:1640–1644

Youssefian S, Nakamura M, Orudgev E, Kondo N (2001) Increased cysteine biosynthesis capacity of transgenic tobacco overexpressing an acetylserine(thiol) lyase modifies plant responses to oxidative stress. Plant Physiol 126:1001–1011

Yu J, Hu S, Wang J, Wong GK-S, Li S et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L ssp indica) Science 296:79–91

Yu JK, Dake TM, Singh S, Benscher D, Li WL, Gill B, Sorrells ME (2004) Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. Genome 47:805–818

Yue SJ, Li H, Li YW, Zhu YF, Guo JK, Liu YJ, Chen Y, Jia X (2008) Generation of transgenic wheat lines with altered expression levels of 1Dx5 high-molecular-weight glutenin subunit by RNA interference. J Cereal Sci 47:153–161

Zhang D, Choi DW, Wanamaker S, Fenton RD, Chin A, Malatrasi M, Turuspekov Y, Walia H, Akhunov ED, Kianian P, Otto C, Simons K, Deal KR, Echenique V, Stamova B, Ross K, Butler GE, Strader L, Verhey SD, Johnson R, Altenbach S, Kothari K, Tanaka C, Shah MM, Laudencia-Chingcuanco D, Han P, Miller RE, Crossman CC, Chao S, Lazo GR, Klueva N, Gustafson JP, Kianian SF, Dubcovsky J, Walker-Simmons MK, Gill KS, Dvorák J, Anderson OD, Sorrells ME, McGuire PE, Qualset CO, Nguyen HT, Close TJ (2004) Construction and evaluation of cDNA libraries for large-scale expressed sequence tag sequencing in wheat (*Triticum aestivum* L). Genetics 168:595–608

Zhao XY, Cheng ZJ, Zhang XS (2006) Overexpression of *TaMADS1*, a *SEPALLATA*-like gene in wheat, causes early flowering and the abnormal development of floral organs in Arabidopsis. Planta 223:698–707

Zhou HB, Li SF, Deng ZY, Wang XP, Chen T, Zhang JS, Chen SY, Ling HQ, Zhang AM, Wang DW, Zhang XQ (2007) Molecular analysis of three new receptor-like kinase genes from hexaploid wheat and evidence for their participation in the wheat hypersensitive response to stripe rust fungus infection. Plant J 52:420–434

Chapter 13
TILLING for Mutations in Model Plants and Crops

Zerihun Tadele, Chikelu MBA, and Bradley J. Till

Abstract A growing world population, changing climate and limiting fossil fuels will provide new pressures on human production of food, medicine, fuels and feed stock in the twenty-first century. Enhanced crop production promises to ameliorate these pressures. Crops can be bred for increased yields of calories, starch, nutrients, natural medicinal compounds, and other important products. Enhanced resistance to biotic and abiotic stresses can be introduced, toxins removed, and industrial qualities such as fibre strength and biofuel per mass can be increased. Induced and natural mutations provide a powerful method for the generation of heritable enhanced traits. While mainly exploited in forward, phenotype driven, approaches, the rapid accumulation of plant genomic sequence information and hypotheses regarding gene function allows the use of mutations in reverse genetic approaches to identify lesions in specific target genes. Such gene-driven approaches promise to speed up the process of creating novel phenotypes, and can enable the generation of phenotypes unobtainable by traditional forward methods. TILLING (Targeting Induced Local Lesions IN Genome) is a high-throughput and low cost reverse genetic method for the discovery of induced mutations. The method has been modified for the identification of natural nucleotide polymorphisms, a process called Ecotilling. The methods are general and have been applied to many species, including a variety of different crops. In this chapter the current status of the TILLING and Ecotilling methods and provide an overview of progress in applying these methods to different plant species, with a focus on work related to food production for developing nations.

Z. Tadele
Institute of Plant Sciences, University of Bern, Switzerland
e-mail: zerihun.tadele@ips.unibe.ch

C. MBA and B.J. Till (✉)
Plant Breeding Unit, FAO/IAEA Agricultural and Biotechnology Laboratory, Seibersdorf, Austria
e-mail: c.mba@iaea.org, b.till@iaea.org

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_13,

13.1 Introduction

Agricultural sustainability can be defined as a state where increases in food production are at least proportional to the rate of population growth. The projected world population by the year 2050 will be 9.2 billion; the bulk of which will be from developing and least developed countries. It is estimated that increases in population will be 4.3%, 61.2% and 156.4% for developed, developing and least developing nations, respectively (United Nations 2007). This suggests that major pressures for crop production will befall nations with the least resources available for crop improvement. Furthermore, many developing nations rely on so-called understudied crops where only minor investments have been made in plant biotechnology. The ability of developed nations to provide food assistance to meet the expected shortfall is in doubt due to increased demands on crops for non-food purposes such as biofuels (von Braun and Pachauri 2006).

Crop productivity is mainly affected by a variety of abiotic and biotic stresses. Approximately 70% of potential yield is lost due these stresses (Gale 2002). The major abiotic stresses that affect food production are drought, salinity and acidity. From the total global arable area, a third is affected by salinity and 40% by acidity (Gale 2002). Biotic factors such as diseases, insects and weeds also contribute to decreased yields. Crop production is increased either by expanding the arable area or using inputs such as improved seed, irrigation and chemicals. According to the Food and Agriculture Organization (FAO), about 80% of future increases in crop production in developing countries are predicted to come from agricultural intensification (FAO 2002). Based on this goal, crop breeders focus towards achieving improved cultivars that produce higher yields and at the same time tolerate to the sub-optimal soil and climatic conditions. By utilizing various breeding techniques, a number of improved cultivars from different species have reached farming communities and contributed to increases in global food production. The work on induced mutation alone led to over 2,000 officially registered crops from 1940 to 2000, of which 85% were the result of gamma- and x-ray mutagenesis (Maluszynski et al. 2000). Induced mutations have had a large impact in transforming the agriculture of the world, particularly in generating crop species having desirable traits (Ahloowalia et al. 2004).

Examples of improved traits with high impact are those which alter the architecture of the plant. Architectural changes include alteration in branching pattern and reduction in plant height. The major achievement of the Green Revolution in 1960s and 1970s was due to the introduction of semi-dwarf cultivars of wheat and rice along with crop production packages such as controlled use of irrigation, fertilizer, herbicide and fungicide. Semi-dwarf varieties have been produced by exploiting natural nucleotide variation and through induced mutations, and have led to tremendous increases in productivity. This led to an annual yield increment of about 80 kg/ha in wheat and rice produced in developing countries between 1975 and 1984 (Conway and Toenniessen 1999). According to the International Food Policy Research Institute, the Green Revolution represented the successful adaptation and transfer of scientific revolution in agriculture (IFPRI 2002). Currently, a number of genes affecting plant height have been identified from major cereal crops including wheat, rice and maize (for review, Wang and Li 2006).

With expanding genomic DNA sequence information from many plant species and increasing knowledge regarding the functional roles of specific genes in traits of agronomic importance, it is now possible to consider creating specific plant traits in a directed manner. One approach is to use transgenes to transfer a single or multiple genes of interest within or across species. Using this approach, scientists have been able to create rice producing provitamin A in the grain (Ye et al. 2000). While very powerful, transgenic approaches have been met with a high level of public disapproval and the use of the methods for food production is currently banned in many countries.

This calls for an alternative, non-transgenic, targeted approach for crop improvement in order to meet the increasing demand in food production. TILLING (Targeting Induced Local Lesions IN Genomes) is a general reverse genetic technique that uses traditional mutagenesis followed by high-throughput mutation discovery to identify deleterious lesions in specific target genes. The technique has proven to be robust and easily applied to a variety of different species including *Arabidopsis*, (Mccallum et al. 2000b; Colbert et al. 2001; Greene et al. 2003; TILL et al. 2003), rice (Sato et al. 2006; Till et al. 2007; Suzuki et al. 2008), maize (Till et al. 2004a), wheat (Slade et al. 2005), barley (Caldwell et al. 2004), soybean (Cooper et al. 2008), pea (Triques et al. 2007), and a number of non-plant species such as *Caenorhabditis elegans*, *Drosophila* and zebrafish (Gilchrist et al. 2006a; Winkler et al. 2005; Wienholds et al. 2003). We describe in this chapter the current status of TILLING in plants and discuss the potential role of TILLING in improving under-studied crops of the developing world. We also describe a related technique known as Ecotilling that is used to investigate natural nucleotide variations in the genes of interest.

13.2 TILLING Method

The TILLING strategy utilizes traditional mutagenesis followed by high throughput mutation discovery (Mccallum et al. 2000b; Colbert et al. 2001). The main steps in TILLING are mutagenesis, the development of a non-chimeric population, preparation of a germplasm stock, DNA extraction and sample pooling, screening the population for induced mutations, and the validation and evaluation of mutants (Fig. 13.1). The methods required for each step can be applied to many species, making the TILLING process broadly applicable. Mutants discovered by TILLING can be used for gene-function studies and can be introduced into breeding programs.

13.2.1 Selecting a Mutagen for TILLING

The choice and dose of mutagen for TILLING will dictate the spectrum and density of mutations to be found in individual plants. The best mutagen can be defined as one that produces the desired density and spectrum of alleles with the least pleiotropic

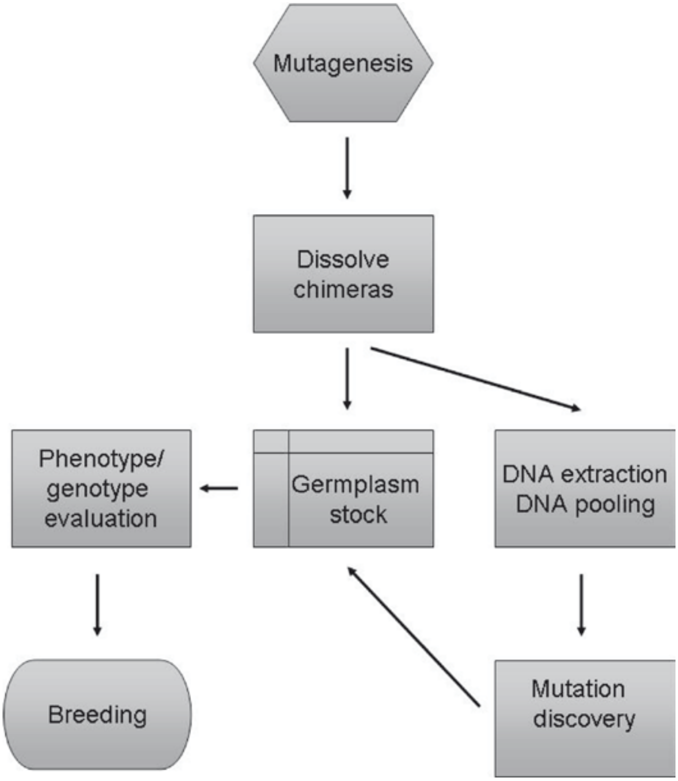


Fig. 13.1 A TILLING pipeline for gene function analysis and developing new crop varieties. A mutagenized population is prepared using a mutagen that primarily causes small lesions (single nucleotide polymorphisms, or insertions/deletions) randomly throughout the genome. Many mutagenic treatments produce a chimeric plant in the first generation. Chimeras are dissolved and a structured population is typically developed. A germplasm stock is prepared for long term storage of mutant lines, and DNA is extracted from each individual mutant. DNAs are pooled and the library of samples is screened for induced mutations in selected regions of target genes. Candidate mutants are removed from the germplasm stock and further characterized genotypically and phenotypically. Individuals or lines exhibiting the desired characteristics can be incorporated into a breeding program

consequences, low material and labor costs, and minimal health risks. Chemical mutagens have been used for many TILLING projects, with ethyl methanesulfonate (EMS) being most common (Table 13.1). Chemical mutagens are attractive for TILLING because protocols for their use in forward genetic screens have previously been developed for many species, the mutagens are readily available, easy to use, and produce predictable heritable nucleotide changes at high densities. Many chemical mutagens, such as EMS, primarily produce single nucleotide changes in the genome (Greene et al. 2003; TILL et al. 2003). Single nucleotide changes are useful because truncation and missense changes are obtained, providing a broad allelic series. Nonsense mutations that create premature stop codons, and splice site

Table 13.1 Spectrum and density of mutations reported in TILLING populations

Organism	Mutagen	Tissue mutag -enized	Mutation discovery method	Population screened	Targets screened	Induced mutation	Spectrum of mutation	Density of mutation (mutations/kb)	Reference
Arabidopsis	EMS[a]	Seed	Mismatch cleavage and fluorescence detection	3000	192	1890	>99% GC–>AT	~1/200	Greene et al. (2003), TILL et al. (2003)
Barley	EMS	Seed	Mismatch cleavage and denaturing HPLC	4600	2	10	70% GC–>AT	~ 1/1000	Caldwell et al. (2004)
Maize	EMS	Pollen	Mismatch cleavage and fluorescence detection	750	11	17	>99% GC–>AT	~1/500	Till et al. (2004a)
Pea	EMS	Seed	Mismatch cleavage and fluorescence detection	3072	5	60	100% GC–>AT	1/669	Triques et al. (2007)
Rice	EMS	Seed	Mismatch cleavage and fluorescence detection	768	10	27	70% GC–>AT	~1/290	Till et al. (2007)
					10	30	11% AT–≥GC		
	Az-MNU[b]						67% GC–>AT	~1/270	
							20% AT–>GC		
Rice	MNU[c]	Seed	Mismatch cleavage and capillary detection	767	4	24	92% GC–>AT	1/135	Suzuki et al. (2008)
Rice	Gamma ray	Seed	Mismatch cleavage and agarose gel detection	2130	25	6	50% Transversion 17% transition–33% small indel[d]	1/6190 kb	Sato et al. (2006)
Soybean	MNU	Seed	Mismatch cleavage and fluorescence detection	529	7	32	~90% GC–>AT	1/140	Cooper et al. (2008)
	EMS			768	7	12	~90% GC–>AT	1/140	
	EMS			768	7	25	~90% GC–>AT	1/250	
	EMS			768	7	47	~75% GC–>AT[e]	1/550	
Wheat	EMS	Seed	Mismatch cleavage and fluorescence detection	1920	7	>200	~99% GC–>AT	1/40 (tetraploid) 1/24 (hexaploid)	Slade et al. (2005)

[a] Ems: ethyl methanesulfonate.
[b] Az-mnu: sodium azide methyl nitrosourea.
[c] Mnu: methyl nitrosourea.
[d] Exact nucleotide changes not reported.
[e] No significant difference in spectrum between this population and other soybean populations.

mutations that inhibit proper intron splicing can knock out gene function providing null alleles. Missense changes can produce variable affects on protein function and are valuable for obtaining phenotypes from essential genes that would cause organism death if completely knocked out. In addition, missense mutations can provide a broad range of phenotypes that can enhance gene function studies, and be useful in breeding. The utility of point mutations in breeding may be best exemplified by the value of the missense mutations causing the semi-dwarf sd1 phenotype in rice that fuelled the Green Revolution (Sasaki et al. 2002).

The spectrum and density of induced mutations identified in TILLING screens varies by mutagen and by species (Table 13.1). In *Arabidopsis*, maize, and wheat, EMS produced > 99% GC:AT changes that could be identified by TILLING (Greene et al. 2003; TILL et al. 2003; Slade et al. 2005). A study of ~2000 EMS induced mutations in *Arabidopsis* showed that mutations were essentially induced randomly throughout the genome (Greene et al. 2003). Random mutagenesis means that mutations can be discovered anywhere in the genome regardless of target size provided the proper balance of population size and mutation density is obtained. With a high density of ~1 mutation per 200 kb in *Arabidopsis*, a relatively small population of 3,000 mutant individuals is typically screened to deliver ~10 alleles per ~1.5 kb target region (http://tilling.fhcrc.org/arab/status.html). Approximately 5% of EMS induced mutations are truncation changes, and 48% missense changes (Greene et al. 2003). A high mutation density reduces the population size needed to deliver the desired number of alleles, and thus reduces the cost of sample preparation and screening. Mutation density is, therefore, a key factor in determining the scale and costs of a TILLING project. For gene function studies and publicly available TILLING services, a strategy has been to achieve a high mutation density to reduce mutation screening costs. A lower mutation density may be desirable for breeding applications.

Other mutagens can be considered for TILLING. For example, mutagenesis of rice seed with sodium azide plus MNU (methyl-nitrosourea), and soybean seed with MNU produced populations with mutation spectra and densities similar to that observed with EMS (Table 13.1). Inducing mutations with physical mutagens has a long history beginning with the work of Muller exposing Drosophila to X-rays (Muller 1928). With Muller's initial report, scientists embarked upon an extensive use of radiation to induce mutations and 30 years later were reporting extensively on induced mutagenesis mediated by ionizing radiation (Bauer 1957; Mac-Key 1956; Singleton 1955; Smith 1958; Konzak 1957; Sparrow 1956; Gaul 1958; Miksche and Shapiro 1963; Hough and Weaver 1959). More recently, Maluszynski et al. (2000), and Ahloowalia et al. (2004) have provided very comprehensive reviews on officially released induced crop mutants, most of which had been induced to mutate using physical agents.

While induced mutations caused by physical mutagens have a long history in both basic and applied research, less is known about the spectrum and density of lesions in crops caused by such treatments. One of the most widely used physical agents is gamma irradiation. A population of rice mutagenized with gamma-irradiation was recently used in TILLING screens, yielding only point mutations and small

deletions (Sato et al. 2006). The density of induced mutations in this population, however, was very low (1 mutation per 6,190 kb). One explanation for the low density is that gamma-irradiation produced larger genomic lesions in this population that went undetected because of the method used to find mutations. Because large lesions would be, on average, more deleterious to gene function than SNPs (Single Nucleotide Polymorphisms), fewer such lesions will be tolerated in a genome. If a mutagenic treatment induces a mixture of large and small lesions, the result would be a lower maximal density of lesions than, for example, what could be obtained with a mutagen causing only SNPs. Studies of gamma and carbon irradiation of *Arabidopsis* pollen showed that small and large deletions and inversions can occur due to mutagenic treatment (Naito et al. 2005). Together, this suggests that gamma-irradiation of plant material can cause a combination of large and small lesions, and that obtainable mutation densities may be lower when compared with chemical mutagenesis. More work is to be done to characterize the spectrum and density of gamma-induced lesions in plant genomes, and the effects of using different doses of mutagen.

Large genomic lesions such as deletions are useful for creating phenotypes, especially when knockouts are desired, including the deletion of tandemly repeated genes, such as those involved in disease response. For example, fast neutron mutagenesis was used to create deletions in *Arabidopsis* and rice that could be detected by a simple PCR based assay (Li et al. 2001). When compared to mutagens causing primarily SNPs, fast neutron mutagenesis is disadvantageous because mutation densities are much lower, and a ~>tenfold larger population is required to ensure the recovery of a deletion in any gene in the genome. Furthermore, the mutations will cause primarily knockouts versus the allelic spectrum of knock-out and missense changes caused by point mutations. However, physical mutagens that create large deletions and inversions may produce a higher frequency of dominant alleles. This may be desirable, particularly when studying polyploid species or vegetatively propagated crops. An additional advantage of physical mutagenesis is that the procedure is easily centralized and plant material is non-toxic and safe to ship after treatment. For example, the FAO/IAEA joint programme provides a gamma-irradiation service that is free to its member states (http://www.iaea.org/OurWork/ST/NA/NAAL/agri/pbu/agriPBUmain.php). When creating mutagenized populations for TILLING, it may be worthwhile considering several different mutagens for a combined strategy for the recovery of both large deletions and SNP mutations. The method of mutation discovery, however, is an important factor in choosing a mutagenesis strategy (see Section 13.2.3).

13.2.2 Selecting Tissue for Mutagenesis

The optimal tissue for large-scale mutagenesis for TILLING can be defined as one that produces non-chimeric plants harbouring a high density of induced mutations and requires the lowest inputs of labor and propagation time. Because most point

mutations causing phenotypes will be recessive, an additional criterion is that mutations can be easily made homozygous. The optimal tissue may therefore change from species to species and depending on the resources and needs of the scientist. Seed mutagenesis has been the choice for many plant species (Table 13.1). For chemical mutagenesis, seeds are soaked in a mutagen for a set period of time, washed to remove the mutagen, and then sown. The first generation, called M_1, is genotypically chimeric, with different parts of the plant having different genotypes because of the multicellular nature of the embryo at the time of mutagenesis (Henikoff and Comai 2003). To obtain heritable mutations, the plant must be propagated to produce a generation that is non-chimeric before DNA can be collected for TILLING screens. When possible, self-fertilization is used. Populations are typically structured using a single seed descent strategy such that one M_2 progeny from the self cross of the M_1 is selected to create a line for TILLING (for example, Till et al. 2003; Caldwell et al. 2004). Tissue is collected from this M_2 individual for later DNA extraction and mutation screening. M_3 seed are collected from a self pollination of the M_2, and this becomes the germplasm stock. Mutations identified in the M_2 sample are either heterozygous or homozygous. If M_2 plants selected for the TILLING population are chosen randomly, a Mendelian ratio of 2:1 heterozygous to homozygous mutations should be observed in the TILLING screen (Greene et al. 2003).

Non-structured, or bulk, populations can also be used for TILLING. As with a structured population, the first non-chimeric generation can be screened. A disadvantage of the bulk strategy is that siblings sharing the same mutations will be screened, increasing the time and cost of mutation discovery. The strategy may therefore be best for projects focusing on specific phenotypes that can easily be identified and sorted to make a phenotype enriched population. The utility of screening a phenotype-enriched population was shown for *Lotus japonicus*. Perry and colleagues created a mutagenized population of *Lotus* that was enriched for nodule and root-specific phenotypes. TILLING screens of 288 plants led to the identification of 6 novel alleles (Perry et al. 2003). The advantage of this strategy is that the size of the enriched population is much smaller than the entire population, saving time and money on mutation screening. The enriched population will be depleted in mutations causing non-enriched phenotypes, and the complete mutagenized population is screened when targeting genes hypothesized to cause other phenotypes.

Efficiencies in the TILLING pipeline can be gained by choosing tissues for mutagenesis that create a non-chimeric individual in the M_1 generation. For maize, pollen has been soaked with EMS and then applied to ears of unmutagenized tester strains (Till et al. 2004a). Each kernel on the resulting ear contains a unique collection of heterozygous mutations. A single kernel defines a line and DNA for mutation screening can be extracted from the M_1 generation.

In vitro methodologies also have great potential for rapidly achieving homozygosity and minimising, if not totally nullifying, the need for the dissociation of chimeras in mutagenic populations. Cell suspension cultures, relying on totipotency of cells, involves the production of cell lines from callus followed by the

regeneration of plantlets through somatic embryogenesis. Typically, this involves the culturing of single cells and small cell aggregates that proliferate and complete a growth cycle while suspended in liquid medium. Since Nickell's (1956) pioneering work with cell suspension cultures of *Phaseolus vulgaris*, reproducible protocols have been produced for other plant species. This ability to culture individual plant cells, from which whole plantlets will arise, therefore permits the treatment of individual cells with mutagens. The plantlets that arise from each treated cell are genetically similar leading to significant gains in time, especially as the need for several cycles of regenerations required to dissociate chimeras are eliminated. Because plants from culture develop via mitosis, induced mutations will remain heterozygous in adult tissues. To uncover recessive phenotypes, mutations must first be made homozygous.

Totipotency is also exploited in the regeneration of doubled haploids (DHs), when the chromosome number of gametic cells, i.e. pollen and egg cells, is doubled prior to regeneration of a plant (Forster et al. 2007). This process could be incorporated into induced mutagenesis by the treatment of these gametic cells prior to regeneration of the doubled haploids. With spontaneous and/or induced doubling of the haploid chromosomes, homozygous individuals are produced, providing the most rapid mechanism for attaining homozygousity with the greatest fidelity (Szarejko and Forster 2007). By facilitating the possibility of targeting either the haploid or doubled haploid cells for mutation treatment, the authors (Szarejko and Forster 2007) posit that a mutation is captured in a homozygous, pure line. These individuals are homozygous for all loci including the mutated segments of the genome being targeted for detection. The savings in time and cost are significant as recessive mutations usually are not detectable until the M_2 or later generations of sexual propagation through self-fertilization. Rapid technological advances have resulted in the availability of reproducible DH protocols for over 250 plant species (Maluszynski et al. 2003) covering most plant genera. Enthusiastic incorporation of DH in induced mutagenesis should therefore be successful and lead to significantly enhanced efficiency in delivery processes. A major disadvantage, however is that unlike self fertilization, where individual progeny are expected to be homozygous for ~25% of alleles, a plant created by DH procedures will be homozygous for all mutations, including deleterious background mutations that can confound phenotypic analysis. Because of this, DH plants will tolerate a lower mutational load, thus mutation densities by this method will be lower, necessitating larger populations to discover sufficient alleles. For vegetatively propagated crops that produce viable pollen, DH may be the only method for creating homozygous alleles.

Where protocols for somatic embryogenesis, through cell lines or friable embryogenic calli, for instance, are not available, plantlets could also be regenerated but at relatively lowered levels of homozygousity and enhanced levels of chimeral sectors through in vitro nodal segments. For obligate vegetatively propagated crops for which microspore cultures followed by chromosome doubling are impracticable, the in vitro nodal segments strategy, while not optimal, is more efficient than using stem cuttings or other vegetative propagules itemised by Micke and Donini (1993) which include tubers, dormant buds; rhizomes; dormant shoots; etc.

If this route is taken, induced crop mutants would be developed, even if genetic constitution, including other mutation events, are unknown (Van Harten 1998) but due consideration must be given to planning strategies for efficient dissociation of chimeras. Where multicellular meristematic tissues have been used as starting materials for the induction of mutations, several cycles of regenerations are required to dissociate chimeras in order to approximate the homohistont state. The optimal number of regenerations to remove chimerism may vary by species, propagation techniques and the type of chimeras (mericlinal, periclinal or sectorial chimeras), and should be determined empirically. For instance, Novak et al. (1990) and Roux (2004) suggested two different numbers of regenerative cycles, 6 and 4, i.e. M_1V_6 and M_1V_4, respectively as the minimum for banana. Also, because mutations are expected to be heterozygous, choosing a mutagenesis treatment that increases the frequency of dominant alleles should be a top priority.

13.2.3 DNA Extraction, Pooling and Mutation Discovery

The DNA extraction method used for TILLING should produce a sufficient yield and quality of genomic DNA that is stable for the duration of the project. These parameters are determined empirically (Till et al. 2006b). Genomic DNA samples are typically prepared from single plants. Samples are normalized to a common concentration and pooled together to reduce the cost of mutation screening. Pooling of up to eight samples has been used for TILLING (for example, Colbert et al. 2001; Till et al. 2004a; Cooper et al. 2008). Several strategies for pooling samples have been used. In a one dimensional pooling strategy, each plant is represented in a single pool of eight, and 768 unique samples (96 pools) can be screened in a single 96 well TILLING assay. Pools with putative mutations are then deconstructed and the eight samples comprising the pool are screened individually (Colbert et al. 2001; Till et al. 2006b). A two-dimensional pooling strategy has also been used (Till et al. 2007; Cooper et al. 2008). In this strategy, individual samples are arrayed in 8 × 8 grids and pools are made by combining samples by columns and by rows (Fig. 13.2). An individual sample is therefore represented in two unique pools in the assay, and a positive mutation produces a signal in two assay lanes. The repetition of samples in the two dimensional strategy means that at a constant sample pooling (i.e. eightfold), only half the number of samples are screened per assay when compared to the one dimensional method (384 versus 768 for 96 well assays). The advantage of the two dimensional strategy is that mutations are validated and individuals identified in a single step, streamlining the process and potentially minimizing false positive and false negative error rates (Till et al. 2006a).

A variety of methods are available for the discovery of unknown nucleotide polymorphisms. For example, traditional Sanger sequencing, denaturing HPLC and enzymatic mismatch cleavage have all been used for mutation discovery in reverse genetic screens of populations treated with chemical mutagens (McCallum et al.

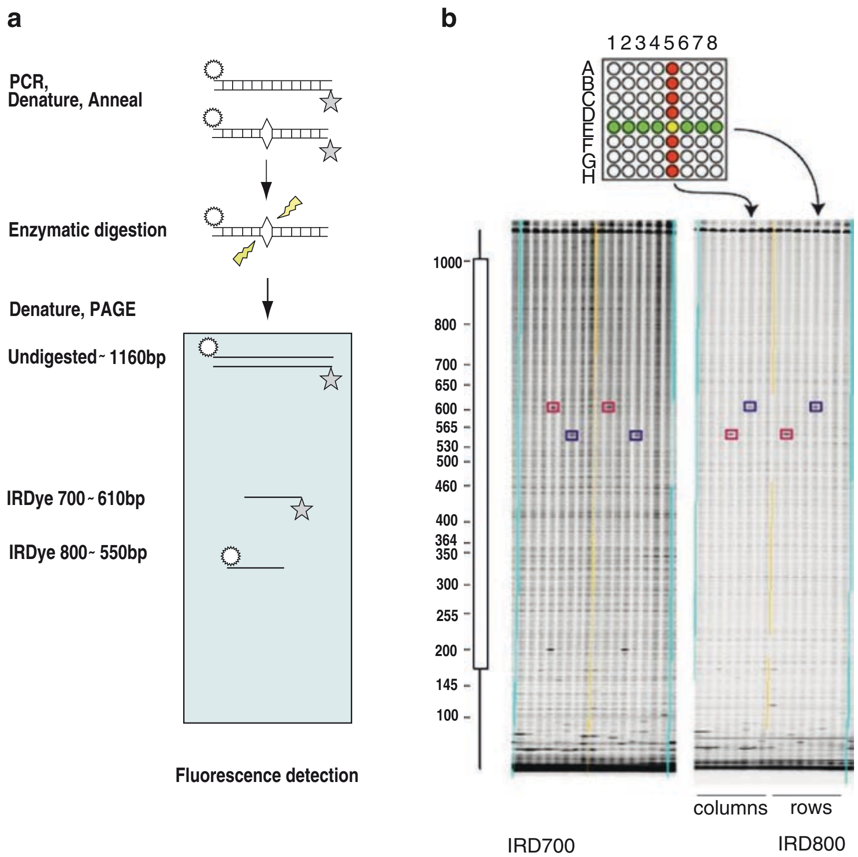


Fig. 13.2 Enzymatic mismatch cleavage and fluorescence detection for mutation discovery in TILLING. Schematic diagram of enzymatic mismatch cleavage (panel A). PCR is performed with primers end-labelled with IRDye700 (*circle*) and IRDye800 (*star*) fluorescent dyes. After PCR, amplicons are denatured and annealed to form heteroduplexed molecules. Mismatched regions are cleaved by treatment with a single-strand specific nuclease, typically an extract containing CEL I. Products are then denatured and size fractionated by denaturing polyacrylamide gel electrophoresis. DNA is visualized by fluorescence detection, typically with a LI-COR DNA analyzer. The sum of the molecular weights of the cleaved products equals the molecular weight of the undigested full length product. Using two different dyes allows the approximate position of the mutation on the PCR amplicon to be estimated. A two dimensional pooling strategy was used to discover chemically induced mutations in rice (panel B). Sixty-four samples are arrayed in an eight by eight grid (*top*). Column pools are prepared by combining all samples in a common column (numbered 1–8). The same is done for row pools (A-H). Samples from column and row pools from a grid are loaded on adjacent lanes of the gel (*bottom*). A positive signal found in a column pool will replicate in a row pool. The lanes provide coordinates for determining the identity of the mutant sample in the grid (E5 in the marked example). Panel B modified from Till et al. 2007

2000a; Colbert et al. 2001; Wienholds et al. 2002; Caldwell et al. 2004). Factors important in choosing a mutation discovery platform include equipment cost, maintenance costs, reagent costs, and reliability and error rates of the assay. Currently,

the most common platform is enzymatic mismatch cleavage using single-strand specific nucleases followed by fluorescence detection (Fig. 13.2, Table 13.1). PCR is performed using a forward primer 5′ end labelled with the IRDye700 dye, and a reverse primer labelled with IRDye800 to produce amplicons of approximately 1.5 kb (Fig. 13.2a). Products are then denatured and annealed to form heteroduplexes between polymorphic amplicons. Heteroduplexes are cleaved with a single-strand specific nuclease and products are size fractionated by denaturing polyacrylamide gel electrophoresis (PAGE). Fluorescent signals are detected using the Li-Cor DNA analyzer (Fig. 13.2b). The molecular weights of the cleaved fragments indicate the position of the nucleotide change within ~10 bp (Greene et al. 2003). The exact nucleotide change is then determined by sequencing. Mung bean nuclease, S1 nuclease, CEL I nuclease, and other members of the S1 family have been shown to work in the TILLING assays (Till et al. 2004b; Triques et al. 2007). The accuracy of using a crude extract of CEL I to detect nucleotide polymorphisms in pools was determined using human SNPs. In a blind study testing discovery efficiencies using previously identified human polymorphisms, a false discovery rate of 2% and false negative error rate of 7% was reported for screening samples in eightfold pools that were arrayed two dimensionally (Till et al. 2006a). The high throughput and low cost nature of this mutation discovery method for TILLING is exemplified by fees charged by TILLING services that operate on a full-cost recovery basis… At the time of writing, users pay USD dollar 2,000 for an allelic series of ~12 induced mutations from ~3,000 mutagenized *Arabidopsis* lines (http://tilling.fhcrc.org/files/user_fees.html). The sensitivity of the mutation discovery method in a production setting is shown by the recovery of the expected 2:1 Mendelian ratio of heterozygous to homozygous mutations in the samples from M_2 *Arabidopsis* plants (Greene et al. 2003). If mutation discovery sensitivity was limiting one would expect to find less heterozygous mutations because they are represented at half the concentration of homozygous polymorphisms in pooled samples.

The TILLING method has been further streamlined through the development of computational tools for assay design and data analysis. The CODDLe (Codons Optimized to Detect Deleterious Lesions) program uses the expected spectrum of mutations for a given mutagen to calculate the ~1.5 kb region of a gene that contains the highest number of potential mutations that could adversely affect gene function (McCallum et al. 2000b, http://www.proweb.org/coddle). Because a high density of mutations can be achieved, sufficient alleles are typically obtained by choosing the best 1.5 kb region of the gene rather than screening the entire gene in overlapping 1.5 kb segments. In addition to counting nonsense and splice site changes, CODDLe uses protein homology to predict the severity of missense changes. The CODDLe output links directly to Primer3 primer design software (Rozen and Skaletsky 2000). Approximately 90% of primers ordered for the *Arabidopsis* TILLING Project service ordered using this system have produced sufficient amplification product for TILLING (http://tilling.fhcrc.org/arab/status.html). Once mutations have been discovered by TILLING, they can be graphically displayed with the PARSESNP program that incorporates the SIFT program to predict if missense mutations are likely to affect protein function (Taylor and

Greene 2003; Ng and Henikoff 2003; http://www.proweb.org/parsesnp). Analysis of Li-Cor gel data for TILLING and Ecotilling is aided by the PC/Mac program Gelbuddy (Zerr and Henikoff 2005; www.gelbuddy.org). Gelbuddy provides molecular weight calibration, automated lane discovery, and automated band discovery. All programs described here are freely available and can be used for additional applications not linked to TILLING or Ecotilling.

Rapid advancements in next generation sequencing platforms hold great promise for increasing the efficiency of mutation discovery for TILLING. Next-generation sequencing can be defined as any number of new technologies that promise to dramatically increase the speed and reduce the cost of DNA sequencing when compared to traditional Sanger sequencing. Technologies include currently commercial pyrosequencing and mass spectroscopy based platforms to the developing field of nanopore sequencing (Hall 2007; Shendure et al. 2008). While whole genome sequencing is not at the time of writing cost competitive for the discovery of the approximately several hundred induced mutations in a highly mutagenized diploid plant in a population of thousands of plants, progress is being made in strategies for selective enrichment of desired targets that drive costs downward. Strategies include pre amplification of selected targets by PCR and genomic selection by microarray hybridization (for example, Albert et al. 2007).

Several large scale TILLING services including the Seattle TILLING Project, the Maize TILLING Project and the Rice TILLING Project, are currently evaluating next generation sequencing technologies for TILLING using the Solexa and ABI SOLiD platforms (Henikoff S, Comai L, personal communication, and http://genome.purdue.edu/maizetilling/). Next generation sequencing technologies can be used to discover all types of lesions, making them suitable for mutation discovery regardless of the choice of mutagen and type of lesion created. In addition, mutation discovery in polyploids may be more efficient using sequencing methodologies that collect data from a single starting molecule because it is not necessary to target specific homeologues (see Section 13.4). The accuracies and costs reported for TILLING using enzymatic mismatch cleavage will provide valuable data as a baseline to evaluate next generation sequencing technologies for TILLING, and can provide direction for future technological advancements. In the near-term, implementing new sequencing technologies will remain expensive, and likely financially feasible only for large scale facilities. Indeed, the high cost of laboratory infrastructure and training for mutation discovery, along with the generality of the methods, suggests that centralized mutation discovery facilities will be the most efficient and cost effective means to support TILLING for the scientific community. However, it is expected that the cost of sequencing technologies will drop dramatically, and early successes by large scale facilities will be valuable for smaller groups interested in TILLING. While new technologies are very exciting, it is important to note that the current state of the art for TILLING mutation discovery, enzymatic mismatch cleavage followed by fluorescence detection, is not a bottleneck in terms of time or cost for the TILLING pipeline as outlined in Fig. 13.1. The major bottleneck lies in mutant characterization, and large efficiencies can be gained by developing rapid and low cost phenotyping procedures. Advances in mutation discovery will, however,

greatly benefit TILLING facilities that provide fee-based mutation discovery, but not phenotyping, services.

13.3 Examples of TILLING Projects

The TILLING method has continued to gain in popularity since its first description in 2000. There are many active TILLING projects; some are at the level of fully operational TILLING services, while others are just at the beginning of platform development for a new species. TILLING has been adapted to over 20 species, and many groups host web sites describing projects and progress (Table 13.2).

13.3.1 High-Throughput Services

TILLING projects can be grouped into two broad categories: internally focused projects aimed at addressing specific biological problems, and service-based projects aimed at providing screening services to one or more research community. The first publicly available TILLING service was the Arabidopsis TILLING Project run by the Seattle TILLING Project (http://tilling.fhcrc.org/). Since its inception in 2001, the Arabidopsis TILLING Project has delivered ~8,000 mutations to the research community (http://tilling.fhcrc.org/arab/status.html). With the success of Arabidopsis, the STP went on to develop TILLING in a variety of different organisms for both large and small scales. The STP initiated pilot projects in maize, rice and Drosophila have been developed into large-scale screening services (http://genome.purdue.edu/maizetilling/, http://tilling.fhcrc.org/fly/Welcome_to_Fly-TILL.html, and http://tilling.ucdavis.edu/index.php/Main_Page,)

Other groups have also developed TILLING services (Table 13.2). For example, the TILLMore facility at the University of Bologna in Italy offers fee for service screening of barley (*Hordeum vulgare* cv Morex). The Scottish Crops Research Institute (SCRI) also offers screening services for EMS induced mutations in barley. The Lotus TILLING facility at John Innes Centre in UK provides screening services for EMS induced mutations in *Lotus japonicus*. The Plant Genomics Research Unit, URGV in France currently offers screening services in tomato (TOMATILL), pea (PETILL), and Rapeseed (RAPTIL) (http://urgv.evry.inra.fr/UTILLdb). An internationally organized Grain Legumes Technology Transfer Platform (GL-TTP) runs a Medicago TILLING platform. GABI-TILL is another large consortium with 13 collaborating institutions in Germany that offer screening services for collaborators. The consortium focuses on crops such as barley, sugar beet and potato. The number of services highlights the demand for induced mutations and reverse genetic screening services in a large number of organisms. Based on this, we expect the number of TILLING services to continue to grow.

Table 13.2 Examples of global TILLING projects in plants

TILLING platform/ project	Host organization	Organisms[a]	URL of the project or platform
Arcadia Biosciences TILLING	Arcadia Biosciences, USA	Wheat, castor, other crops	http://www.arcadiabio.com/toolbox.php
Barley TILLING	SCRI, Scotland	Barley	http://www.scri.ac.uk/research/genetics/BarleyTILLING
CAN-TILL	University of British Columbia, Canada	Arabidopsis, brassica, *C.elegans*	http://www.botany.ubc.ca/can-till/
GABI-TILL Project	GABI Consortia, Germany	*Arabidopsis*, barley, sugar beet, potato	http://www.gabi-till.de/main/main/home.html
Lotus TILLING	John Innes Centre, UK	Lotus	http://www.lotusjaponicus.org/tillingpages/homepage.htm
Maize TILLING Project	Purdue University, USA	Maize	http://genome.purdue.edu/maizetilling/
MBGP TILLING	Multinational Consortia	*Brasicca napus, B. oleracea, B. rapa*	http://www.jic.ac.uk/staff/lars-ostergaard/tilling/tilling.htm
Medicago TILLING platform	Grain Legumes Technology Platform, European Consortia	Medicago	http://www.gl-ttp.com/products_services/technical_services/genomic_resources_from_glip/functional_genomics/medicago_tilling_platform/
Peanut TILLING	Georgia Peanut Commission	Peanut	http://www.gapeanuts.com/growerinfo/research/research2006.asp
Rice TILLING platform	Nat. Inst. Of Genetics, Japan	Rice	http://www.nig.ac.jp/section/kurata/kurata-e.html
Seattle TILLING Project	Seattle, USA	*Arabidopsis*, Drosophila (services). Pilot projects in rice, soybean, and other organisms	http://tilling.fhcrc.org/
Soybean Mutation Project	Southern Illinois University, USA	Soybean	http://www.soybeantilling.org/
TILLmore	DiSTA, Bologna, Italy	Barley morex	http://www.agrsci.unibo.it/~salvi/tillmore/index.htm

(continued)

Table 13.2 (continued)

TILLING platform/ project	Host organization	Organisms[a]	URL of the project or platform
University of California-Davis TILLING Project	UC Davis, USA	Rice, tomato, wheat	http://tilling.ucdavis.edu/index.php/Main_Page
URGV TILLING Project	URGV, France	Pea, rapeseed, tomato	http://www.versailles.inra.fr/urgv/analysis-cropFunctionalGen.htm
USDA Bean TILLING Project	USDA	Beans	http://www.ars.usda.gov/research/publications/publications.htm?SEQ_NO_115=197922
USDA Sorghum TILLING Project	USDA	Sorghum	http://www.ars.usda.gov/research/publications/publications.htm?SEQ_NO_115=180731

[a]List of organisms may be incomplete.

13.3.2 Other TILLING Projects

Small scale pilot projects are useful to evaluate the efficacy of a TILLING strategy (choice and dose of mutagen, method for identifying mutation, etc.), and have led to a large number of original research articles (for example, Table 13.1). This indicates that TILLING is a powerful technique to investigate genetic alterations in organisms with different genome size and ploidy level. For example, the hexaploid wheat with genome size of 1.6×10^{10} bp (over 3.5 times more than rice) was successfully used in a TILLING project to create a variety producing low amounts of amylose (Slade et al. 2005).

In addition to public screening services, large-scale projects devoted to specific biological questions have been developed. For example, the Centre for Novel Agricultural Products (CNAP) at the University of York has recently initiated a project to obtain high yielding cultivars of *Artemisia annua*, the source of the anti-malarial component artemisinin, using the TILLING technique (CNAP 2006). Artemisinin-based therapies (ACTs) remain one of the most effective treatments against different species of *Plasmodium*, the causal agents of malaria (Enserink 2007). Because of its high demand, the drug is becoming more expensive and is not affordable in poorer countries where the epidemics of malaria are high. Thus TILLING could have a major impact in the production of medicines, exemplifying the utility of TILLING for crop improvement for non-food production. TILLING can also be envisioned as a tool for the production of superior crops for other non-food uses such as biofuel production.

13.4 Challenges for Crops

Success with organisms such as maize, rice, soybean and wheat show the utility of the TILLING method for crops. Some crops face unique challenges for TILLING. Polyploidy may at first seem an obstacle, but work with wheat shows that TILLING provides a targeted method for developing phenotypes that may not be obtainable when using forward genetic strategies (Slade et al. 2005). With TILLING in polyploids, severe alleles in different homeologues can be discovered independently, and then combined by cross-fertilization to provide the desired phenotype. The main technical challenge for TILLING in polyploids is the need to PCR amplify only one gene target (homeologue) per TILLING assay using enzymatic mismatch cleavage for mutation discovery. In theory, multiple gene targets could be amplified and screened at the same time, however in practice it has been observed that this approach leads to the failure to detect mutations (Cooper et al. 2008). The same issue arises when screening closely related genes in diploids. Amplification of a single gene target can be accomplished by careful primer design (Slade et al. 2005), or by removing the undesired gene target by restriction digestion of the genomic DNA prior to PCR (Cooper et al. 2008). Lack of genome sequence data can also be considered a major

challenge. While inefficient, primers for TILLING can be developed using cDNA or EST sequence data. PCR products from primers are then tested to determine intron location and size. This is the strategy used for pilot projects at the STP, and by the Maize TILLING Project (http://genome.purdue.edu/maizetilling/index.htm).

One concern in implementing TILLING in breeding will be the mutational load, or the number of background mutations in a particular plant. Background mutations can modify, mask, or enhance the desired phenotype, or can produce an unlinked and unwanted phenotype. Knowing the mutation density from TILLING assays allows an estimation of the total number of mutations in a single plant. This information can be used when deciding the best course of action to deal with background mutations for gene function studies and breeding. For example, using the mutation frequency for *Arabidopsis* mutants, it was estimated that through genotype segregation analysis of the progeny from a heterozygous mutation identified in the M_2 generation, a phenotype would be misattributed at a frequency of ~0.0005 (Henikoff and Comai 2003). Furthermore, using known recombination frequencies, the number of backcrosses required to remove the desired number of background mutations can be calculated. Because the majority of point mutations causing phenotypes are likely recessive, and background point mutations will be distributed randomly from plant to plant, a rapid strategy to remove the effects of background mutations is by crossing two strong alleles together as is done in a complementation or allelism test. This results in all background mutations becoming heterozygous. Because the probability that two plants randomly accumulated background mutations deleterious to the same gene is vanishingly small, the phenotypes resulting from the complementation cross can be attributed to the target gene with high confidence. This strategy is suggested by the *Arabidopsis* TILLING project for gene function analysis (http://tilling.fhcrc.org/files/FAQ.html).

For vegetatively propagated crops, the challenges of developing a suitable TILLING population mainly lie in the choice of mutagen, tissue to be mutagenized and the ability to either obtain a sufficient density of dominant mutations, or to create double haploid plants. Work with vegetatively propagated crops is just beginning and the information regarding efficient approaches will have to be gathered through trial and error. Where resources for in vitro techniques are limited, small TILLING populations can be generated and then rapidly screened, and only plantlets with interesting mutations maintained for further analysis. This live population strategy has been previously used for Drosophila and Zebrafish (Bentley et al. 2000; Winkler et al. 2005; Wienholds et al. 2003).

13.5 The Role of TILLING in Orphan Crops

Understudied crops also known as orphan, underutilized, lost, or disadvantaged crops, play major roles in the economy of developing countries (Naylor et al. 2004). They provide income for subsistence farmers and serve as staple food for largely low income consumers. Because of genetic diversity and localized selective

pressures not found in major crop monocultures, these under researched crops can perform better than major crops of the world under extreme soil and climatic conditions prevalent in developing world, particularly in Africa (Ketema 1997; Nelson et al. 2004; Williams and Haq 2002). The robust phenotypes and underlying genetic information will likely become valuable as crop performance is affected by climate change.

13.5.1 The Need to Improve Orphan Crops

As their name suggests, despite their huge importance, understudied crops have so far received little attention from the scientific community. Due to lack of genetic improvement, these crops produce inferior yield in terms of quality and quantity. Moreover, some of these crops produce a variety of toxins which are hazardous to humans if consumed before post-harvest processing (Getahun et al. 2003; Vetter 2000).

In general the major bottlenecks affecting the productivity of under-researched crops are genetic traits such as low yield (for example, in tef (*Eragrostis tef*), finger millet *(Eleusine coracana*)), poor in nutrition cassava (*Manihot esculenta*), enset (*Ensete ventricosum*)), and production of toxic substances (cassava and grass pea (*Lathyrus sativus*)). Environmental factors such as drought, soil acidity and salinity, pests, diseases and weeds also contribute for large loss in quality and quantity of yield. Hence, an agricultural revolution is needed to increase food production of understudied crops in order to feed the ever increasing population in the developing world.

13.5.2 TILLING Projects in Understudied Crops

The advancement of genomic techniques and information is making the process of moving an organism from "understudied" to "well-studied" easier than ever before. Genome sequencing projects are in progress or proposed for several understudied plant species (http://www.ncbi.nlm.nih.gov/sites/entrez?db=genomeprj), and collections of cDNA and EST sequences continue to grow, enabling comparative genomics studies that can provide predictions regarding gene function. With sequence information and gene function predictions, TILLING becomes ideally suited for understudied species that lack many of the tools common to well studied species. Because it is a non-transgenic method, it can be easily incorporated into breeding programs without regulatory restrictions. In this section, we describe examples of progress in TILLING two understudied crops and describe unique challenges for these species.

Cassava (manioc; *Manihot esculenta* Crantz), a herbaceous plant with starchy tuberous storage roots and a major food security crop especially in sub-Saharan

Africa has an annual cultivation acreage of about 16 million hectares. The edible storage roots constitute the main source of dietary calories for about 500 million people in Africa, Asia and Latin America (http://www.ciat.cgiar.org/ciatinfocus/cassava.htm). The estimated production for 2007 was 200 million tons (http://www.fao.org/docrep/010/ah864e/ah864e06.htm#32). In addition to starch, the storage roots contain significant amounts of calcium, phosphorus and vitamin C (Nutrient Data Laboratory, USDA). The cassava plant is hardy, tolerant to drought and capable of producing far more meaningful yields under low input agriculture and on poor soils than most major crops. Cassava production is constrained by myriad problems, with the most significant being biotic stresses (principally bacterial and viral diseases). Additionally, the high contents of the anti-nutritional factor, cyanogenic glucosides (CG) a deadly poison, coupled with low protein contents make the crop unsuitable as a staple. Another bottleneck to large-scale cultivation of cassava is the rapid post-harvest physiological deterioration – the production of scopoletin as a wound response mechanism which in turn darkens the roots within 48 h of harvest (Ceballos et al. 2004; Vetter 2000). Being the cheapest source of starch, the crop has potentials for replacing other starch sources for varied industries including pharmaceutical, textile, paper and bioethanol applications.

The Joint FAO/IAEA Programme has therefore initiated a TILLING project in cassava as a means for directly querying the putative mutants for lesions in target genes before further field trialling. The mutants developed jointly with International Centre for Tropical Agriculture (CIAT) are an important resource for developing the cassava TILLING and Ecotilling platforms. The cassava genome is currently being sequenced and the imminent increase in sequence information will lead to a routine integration of TILLING in cassava induced mutagenesis as a crop improvement strategy. Conversely, induced mutagenesis and efficient reverse genetics strategies such as TILLING will contribute significantly to a rapid use of the bourgeoning sequence information in functional genomics studies in cassava.

Parallel to setting up the cassava TILLING platform, efforts are also being directed at mitigating the effects of chimerism in cassava mutants through the development of validated protocols for the integration of somatic embryogenesis in cassava mutagenesis. Protocols for profusely regenerating plantlets from friable embryogenic callus are available and current efforts target their adaptation to induced mutations (using both physical and chemical mutagens). Being able to eliminate chimeras, regenerating plantlets from a single mutagenised cell, will greatly enhance the predictive value of lesions that are identified through TILLING.

Tef (*Eragrostis tef* (Zucc.) Trotter) is grown annually on over 2.5 million hectares of land mainly in Ethiopia. The plant adapts to diverse climatic and soil conditions and grows better than other cereals both under drought and water-logged conditions (Ketema 1997). Unlike other cereals, the seeds of tef can be stored easily without losing viability under local storage conditions, since it is not attacked by storage pests (Ketema 1997). Tef is free of gluten hence safe for people with severe allergies to wheat gluten (NRC 1996; Spaenij-Dekking et al. 2005). Compared to other cereals, however, the average seed yield from tef is one of the lowest. Lodging is the major constraint to increasing the yield of tef (Ketema 1997).

The tef TILLING project based at the University of Bern in Switzerland is recently initiated with financial support from Syngenta Foundation for Sustainable Agriculture and University of Bern, and scientific collaboration with University of Georgia, FAO/IAEA Programme, and Ethiopian Institute of Agricultural Research. The main goal of the project is to obtain semi-dwarf tetraploid tef lines which are resistant to crop lodging. Since tef has a tall and tender stem, it is susceptible to damage by wind and rain. In addition, when the optimum amount of fertilizer is applied to increase the yield, a high incidence of lodging occurs. As a consequence, the yield from the crop is severely reduced in terms of total grain yield and quality. In general, the yield loss due to lodging is estimated to about 30% for tef. So far, the project has generated over 4,000 M_2 mutagenized lines to be utilized in TILLING.

The genes to be investigated are selected based on the information from other cereals. The dwarf plants of wheat that tremendously increased the yield during the Green Revolution in 1960s and 1970s contain the mutated Reduced height-1 (Rht-B1 and Rht-D1) gene (Peng et al. 1999). The commercially popular rice cultivar known as *semi-dwarf* (*sd-1*), is also defective in a gene involved in gibberellin biosynthesis (Spielmeyer et al. 2002). In addition, the maize *brachytic2* (*br2*) mutants and its ortholog in sorghum *dwarf3* (*dw3*) are also characterized by compact lower stalk internodes (Multani et al. 2003). The height reduction in these plants results from the loss of a P-glycoprotein (PGP) that modulates polar auxin transport in maize stalk (Multani et al. 2003). The Tef TILLING Project will identify from tef several of the orthologous genes indicated above and use the sequence information to screen the mutagenized population. The ability to create dwarf tef by TILLING should have a positive impact on yield when combined with optimized fertilizer use.

13.6 Ecotilling

The same methods that were developed for high throughput TILLING can be applied for the discovery of natural nucleotide polymorpshisms in populations. Proof of principle experiments were performed with 196 *Arabidopsis* accessions known as ecotypes, where the name Ecotilling derives. Screening revealed that multiple polymorphisms, including SNPs, indels (insertion and deletions) and variation in microsatellite repeat number could be efficiently discovered within a single amplicon, contrary to what might be expected from nuclease cleavage of end-labeled DNA (Comai et al. 2004). It was hypothesized that the ability to observe multiple cleaved fragments from end labelled substrates was due to incomplete digestion by CEL I on any one duplexed DNA molecule in the population of molecules being digested. A total of 55 distinct haplotypes were discovered in five ~1 kb *Arabidopsis* gene fragments. To unambiguously assign haplotypes to individual plants, samples were not pooled prior to screening. To uncover homozygous polymorphisms between samples, and equal concentration of DNA from the

sequenced Columbia accession was added to each sample. Screening therefore revealed unique polymorphisms between the test sample and the known sequence of the Columbia reference. Based on comparisons with data collected by Sanger sequencing, it was concluded that low false positive and false negative error rates were associated with Ecotilling.

The accuracy of the Ecotilling method was further explored in a blind study for human SNPs in a collection of samples previously characterized by resequencing (Till et al. 2006a). Five ~ 1.5 kb gene targets were selected for Ecotilling that overlapped regions in genes that were part of a Sanger-based resequencing SNP discovery effort by another group. Ninety samples were in common between the resequencing and Ecotilling data sets. A 4% false discovery and 2% false negative error negative rate was reported for screening unpooled samples. Samples pooled eightfold were screened to increase the efficiency of discovering rare SNPs that may be involved in human disease, and 2% false discovery and 7% false negative error rates were reported. Importantly, new rare alleles were discovered by Ecotilling that went unreported in the resequencing data, suggesting that the high sensitivity of Ecotilling makes it a useful strategy for discovering rare alleles such as those associated with cancer. Using estimates based on cost-recovery fees from large-scale TILLING services, it was estimated that a large scale effort to identify rare cancer mutations by Ecotilling would be ~50-fold less expensive than by using traditional Sanger sequencing.

One important difference between TILLING and Ecotilling is the amount of data that is produced. For a well mutagenzied diploid population, one can expect to find approximately four induced mutations when screening a 1.5 kb region in 768 individuals. The Ecotilling work in *Arabidopsis* and humans revealed hundreds of polymorphisms when screening 96 samples with a 1.5 kb gene target. With an informatics load increase approaching two orders of magnitude, the task of identifying and managing polymorphic bands in Ecotilling data can become a bottleneck. To aid in the analysis of Ecotilling gels, the GelBuddy program was developed for the analysis of Li-Cor gel images (Zerr and Henikoff 2005). The freely available program provides automated lane identification and molecular weight calibration, plus both manual and automated polymorphism band discovery (Till et al. 2006a, www.gelbuddy.org). Importantly, automated band detection discovered a few polymorphisms that were overlooked during manual analysis. While the overall accuracy was lower for automated band detection than by an expertly trained human, the method is objective and can be combined with manual editing to achieve low error rates and reduce data analysis labor.

As with TILLING, the use of the Ecotilling method continues to grow as it is applied to new organisms. Indeed, natural populations may be the best or only resource to study and exploit nucleotide diversity in species where mutagenesis is difficult or impossible. For example, Ecotilling was used to characterize nucleotide variation in 41 accessions of western black cottonwood (*Populus trichocarpa*), to estimate linkage disequilbrium, heterozygosity and nucleotide diversity (Gilchrist et al. 2006b). At the FAO/IAEA joint programme, Ecotilling is being applied to accessions of banana, cassava, and rice with the hopes that alleles important for

biotic and abiotic stress response can be identified. This work is progressing as mutagenized populations for TILLING are being developed, and so data gathered in Ecotilling can be used when choosing targets to screen for induced mutations (Till, BJ, Jankowicz-Cieslak, J, Nakitandwe, J, Bado, S, Afza, R, and Mba C, unpublished). Nieto and colleagues used Ecotilling to screen for candidate polymorphisms in melon associated with resistance to Melon necrotic spot virus (Nieto et al. 2007). A modification of the Ecotilling method using CEL I double strand cutting and an agarose gel readout platform was used to characterize salmon (Garvin and Gharrett 2007). The ability to incorporate Ecotilling into a production screening pipeline is exemplified by the Maize Tilling Project, that now includes screening 48 maize accessions with standard TILLING screens (http://genome.purdue.edu/maizetilling/EcoTILLING.htm). The characterization and exploitation of natural nucleotide diversity will undoubtedly play an important role in crop improvement in the twenty-first century.

13.7 Conclusions

In less than a decade, TILLING has moved from a proof of concept to a well accepted reverse genetic method that has been applied to over 20 different species. Large-scale TILLING services have delivered thousands of induced mutations to the international research community. Ecotilling has also grown in popularity. With increasing pressures on crop productivity expected in the twenty-first century, we predict the use of induced and natural mutation to elucidate gene function and to enhance phenotypes will continue to gain in importance. TILLING and Ecotilling can provide enhanced efficiencies to breeding pipelines and will therefore be useful in meeting expected demands. The advancement of new mutation discovery techniques should provide further improvements to the TILLING and Ecotilling processes.

Acknowledgements ZT is grateful to Syngenta Foundation for Sustainable Agriculture and University of Bern for financial support to the Tef Biotechnology Project. The Food and Agriculture Organization of the United Nations and the International Atomic Energy Agency through their Joint FAO/IAEA Programme of Nuclear Techniques in Food and Agriculture provide generous support for induced crop mutagenesis activities; this makes the work of BT and CM possible at the Agency's laboratories in Seibersdorf, Austria.

References

Ahloowalia BS, Maluszynski M, Nichterlein K (2004) Global impact of mutation derived varieties. Euphytica 135:187–204

Albert TJ, Molla MN, Muzny DM et al. (2007) Direct selection of human genomic loci by microarray hybridization. Nat Methods 4:903–905

Bauer R (1957) The induction of vegetative mutations in *Ribes nigrum*. Hereditas 43:323–337

Bentley A, MacLennan B, Calvo J et al. (2000) Targeted recovery of mutations in Drosophila. Genetics 156:1169–1173

Caldwell DG, McCallum N, Shaw P et al. (2004) A structured mutant population for forward and reverse genetics in barley (*Hordeum vulgare* L.). Plant J. 40:143–150
Ceballos H, Iglesias CA, Pérez JC et al. (2004) Cassava breeding: opportunities and challenges. Plant Mol Biol 56:503–516
CNAP, Centre for Novel Agricultural Products (2006) CNAP Artemisia Project. Press pack. The University of York, UK
Colbert T, Till BJ, Tompa R et al. (2001) High-throughput screening for induced point mutations Plant Physiol 126:480–484
Comai L Young K, Reynolds SH et al. (2004) Efficient discovery of DNA polymorphisms in natural populations by ecotilling. Plant J 37:778–86
Conway G, Toenniessen G (1999) Feeding the world in the twenty-first century. Nature 402:C55–C58
Cooper JL, Till BJ, Laport RG et al. (2008) TILLING to detect induced mutations in soybean. BMC Plant Biol 8:9
Enserink M (2007) Malaria treatments: ACT two. Science 318:560–563
FAO, Food and Agricultural Organization (2002) World agriculture: towards 2015/2030. FAO, Rome
Forster BP, Heberle-Bors E, Kasha KJ et al. (2007) The resurgence of haploids in higher plants. Trends Plant Sci 12:368–375
Gale M (2002) Applications of molecular biology and genomics to geneticenhancement of crop tolerance to abiotic stress – a discussion document. Food and Agriculture Organization of the United Nations, Rome
Garvin MR, Gharrett AJ (2007) DEco-TILLING: an inexpensive method for single nucleotide polymorphism discovery that reduces ascertainment bias. Mol Ecol Notes 7:735–746
Gaul H (1958) Present aspects of induced mutations in plant breeding. Euphytica 7:275–289
Getahun H, Lambein F, Vanhoorne M et al. (2003) Food-aid cereals to reduce neurolathyrism related to grasspea preparations during famine. Lancet 362:1808–1810
Gilchrist EJ, O'Neil NJ, Rose AM et al. (2006a) TILLING is an effective reverse genetics technique for *Caenorhabditis elegans*. BMC Genomics 7:262
Gilchrist EJ, Haughn GW, Ying CC et al. (2006b) Use of Ecotilling as an efficient SNP discovery tool to survey genetic variation in wild populations of *Populus trichocarpa*. Mol Ecol 15:1367–1378
Greene EA, Codomo CA, Taylor NE et al. (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. Genetics 164:731–740
Hall N (2007) Advanced sequencing technologies and their wider impact in microbiology. J Expt Biol 209:1518–1525
Henikoff S, Comai L (2003) Single-nucleotide mutations for plant functional genomics. Annu Rev Plant Biol 54:375–401
Hough LF, Weaver GM (1959) Irradiation as an aid in fruit variety improvement: I. Mutations in the Peach. J Hered 50:59–62
IFPRI, International Food Policy Research Institute (2002) Green Revolution: curse or blessing? IFPRI, Washington DC
Ketema S (1997) Tef [(*Eragrostis tef* (Zucc.)Trotter]. Promoting the conservation and use of underutilized and neglected crops. 12. Institute of Plant Genetics and Crop Plant Research, Gatersleben/International Plant Genetic Resources institute, Rome, Italy. 50 pp
Konzak CF (1957) Genetic effects of radiation on higher plants. Q Rev Biol 32:27–45
Li X, Song Y, Century K et al. (2001) A fast neutron deletion mutagenesis-based reverse genetics system for plants. Plant J 27:235–242
Mac-Key J (1956) Mutation Breeding in Europe. Brookhaven Symp Biol 9:141–156
Maluszynski M, Nichterlein K, van Zanten L et al. (2000) Officially released mutant varieties – the FAO/IAEA Database. Mut Breed Rev 12:1–84
Maluszynski M, Kasha KJ, Szarejko I (2003) Published doubled haploid protocols in plant species. In: Maluszynski M, Kasha KJ, Forster BP, Szarejko I (eds) Doubled hapolid production in crop plants: a manual. Kluwer Academic, Dordrecht, The Netherlands, pp 309–335

McCallum CM, Comai L, Greene EA et al. (2000a) Targeted screening for induced mutations. Nat Biotechnol 18:455–457

McCallum CM, Comai L, Greene EA et al. (2000b) Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. Plant Physiol 123:439–442

Micke A, Donini B (1993) Induced Mutations. In: Hayward MD, Bosemark NO, Romagosa I (eds) Plant Breeding, Principles and Prospects. Chapman and Hall, London, UK, pp 52–62

Miksche JP, Shapiro S (1963) Use of neutron irradiations in the Brookhaven Mutations Program. Technical Report. From International Atomic Energy Agency Symposium on the Biological Effects of Neutron Irradiations, Upton, NY

Muller HJ (1928) The measurement of gene mutation rate in Drosophila, its high variability, and its dependence upon temperature. Genetics 13:279–357

Multani DBS, Briggs SP, Chamberlin MA et al. (2003) Loss of an MDR Transporter in Compact Stalks of Maize *br2* and Sorghum *dw3* Mutants. Science 302:81–84

Naito K, Kusaba M, Shikazono N et al. (2005) Transmissible and nontransmissible mutations induced by irradiating *Arabidopsis thaliana* pollen with gamma-rays and carbon ions. Genetics 169:881–889

Naylor RL, Falcon WP, Goodman RM et al. (2004) Biotechnology in the developing world: a case for increased investments in orphan crops. Food Policy 29:15–44

Nelson RJ, Rosamond L, Naylor RL et al. (2004) The role of genomics research in improvement of “orphan” crops. Crop Sci 44:1901–1904

Nickell LG (1956) The continuous submerged cultivation of plant tissue as single cells. Proc Nat Acad Sci U S A 42:848–850

Nieto C, Piron F, Dalmais M et al. (2007) EcoTILLING for the identification of allelic variants of melon eIF4E, a factor that controls virus susceptibility. BMC Plant Biol 7:34

Ng PC, Henikoff S (2003) SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 31:3812–3814

Novak FJ, Afza R, Van Duren M et al. (1990) Mutation induction by gamma irradiation of *in vitro* cultured shoot-tips of banana and plantain (*Musa* cvs). Top Agric (Trinidad) 60(1):21–28

NRC (National Research Council) (1996) Tef. In: Lost crops of Africa.Vol. I.: Grains. National Academy of Press, Washington DC. Nutrient Data Laboratory, United States Department of Agriculture pp 215–235. http://www.nutritiondata.com/facts-C00001-01c20cX.html#nutrients-per-serving. Accessed May 20, 2008

Peng J, Richards DE, Hartley NM et al. (1999) Green revolution genes encode mutant gibberellin response modulators. Nature 400:258–261

Perry JA, Wang TL, Welham TJ et al. (2003) A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume *Lotus japonicus*. Plant Physiol. 131:866–871

Roux NS (2004) Mutation Induction in *Musa*. In: Jain SM, Swennen, R (eds) Banana improvement: cellular, molecular biology and induced mutations. Science Publishers Inc., Enfield, NH, USA, pp 23–32

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386

Sasaki A, Ashikari M, Ueguchi-Tanaka M et al. (2002) Green revolution: a mutant gibberellin-synthesis gene in rice. Nature 416:701–702

Sato Y, Shirasawa K, Takahashi Y, Nishimura M, Nishio T (2006) Mutant selection from progeny of gamma-ray-irradiated rice by DNA heteroduplex cleavage using Brassica Petiole extract. Breed Science 56:179–183

Shendure JA, Porreca GJ, Church GM (2008) Overview of DNA sequencing strategies. Curr Protoc Mol Biol Chapter 7:Unit 7.1

Singleton WR (1955) The contribution of radiation genetics to agriculture. Agronomy J 47:113–117

Slade AJ, Fuerstenberg SI, Loeffler D et al. (2005) A reverse genetic, Nontransgenic approach to wheat crop improvement by TILLING. Nat Biotechnol. 23:75–81

Smith HH (1958) Radiation in the production of useful mutations. Bot Rev 24:1–24

Spaenij-Dekking L, KooyWinkelaar Y, Koning F (2005) The Ethiopian cereal tef in celiac disease. New Engl J Med 353:1748–1749

Sparrow AH (1956) Cytological changes induced by ionizing radiations and their possible relation to the production of useful mutations in plants. Work Conference on Radiation Induced Mutations. Biology Department, Brookhaven National Laboratory, Upton, New York, pp 76–113

Spielmeyer W, Ellis MH, Chandler PM (2002) Semidwarf (sd-1), "green revolution"rice, contains a defective gibberellin 20-oxidase gene. Proc Nat Acad Sci U S A 99:9043–9048

Suzuki T, Eiguchi M, Kumamaru T, Satoh H, Matsusaka H, Moriguchi K, Nagato Y, Kurata N (2008) MNU-induced mutant pools and high performance TILLING enable finding of any gene mutation in rice. Mol Genet Genomics 279:213–223

Szarejko I, Forster BP (2007) Doubled haploidy and induced mutation. Euphytica158:359–370

Taylor NE, Greene EA (2003) PARSESNP: A tool for the analysis of nucleotide polymorphisms. Nucleic Acids Res 31:3808–3811

Till BJ, Reynolds SH, Greene EA et al. (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. Genome Res 13:524–530

Till BJ, Reynolds SH, Weil C et al. (2004a) Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biol 4:12

Till BJ, Burtner C, Comai L et al. (2004b) Mismatch cleavage by single-strand specific nucleases. Nucleic Acids Res 32:2632–2641

Till BJ, Zerr T, Bowers E et al. (2006a) High-throughput discovery of rare human nucleotide polymorphisms by Ecotilling. Nucleic Acids Res 34:e99

Till BJ, Zerr T, Comai L et al. (2006b) A protocol for TILLING and Ecotilling in plants and animals. Nat Protoc 1:2465–2477

Till BJ, Cooper J, Tai TH et al. (2007) Discovery of chemically induced mutations in rice by TILLING. BMC Plant Biol 7:19

Triques K, Sturbois B, Gallais S et al. (2007) Characterization of *Arabidopsis thaliana* mismatch specific endonucleases: application to mutation discovery by TILLING in pea. Plant J 51:1116–1125

United Nations (2007) World population prospects, the 2006 revision. United Nations Department of Economic and Social Affairs, New York

Van Harten AM (1998) Mutation Breeding: Theory and Practical Applications. Cambridge University Press, Cambridge, UK, pp 353

Vetter J (2000) Plant cyanogenic glycosides. Toxocon 38:11–36

von Braun J, Pachauri RK (2006) The Promises and Challenges of Biofuels for the Poor in Developing Countries. International Food Policy Institute, Washington DC

Wang Y, Li J (2006) Genes controlling plant architecture. Curr Opin Biotechnol 17:123–129

Wienholds E, Schulte-Merker S et al. (2002) Target-selected inactivation of the zebrafish rag1 gene. Science 297:99–102

Wienholds E, van Eeden F, Kosters M et al. (2003) Efficient target-selected mutagenesis in zebrafish. Genome Res 13:2700–2707

Williams JT, Haq N (2002) Global research on underutilized crops. An assessment of current activities and proposals for enhanced cooperation. ICUC, Southampton, UK

Winkler S, Schwabedissen A, Backasch D et al. (2005) Target-selected mutant screen by TILLING in Drosophila. Genome Res15:718–723

Ye X, Al-Babili S, Klöti A et al. (2000) Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science 287:303–305

Zerr T, Henikoff S (2005) Automated band mapping in electrophoretic gel images using background information. Nucleic Acids Res 33:2806–2812

Chapter 14
Microarray Analysis for Studying the Abiotic Stress Responses in Plants

Motoaki Seki, Masanori Okamoto, Akihiro Matsui, Jong-Myong Kim, Yukio Kurihara, Junko Ishida, Taeko Morosawa, Makiko Kawashima, Taiko Kim To, and Kazuo Shinozaki

Abstract Plants respond and adapt to drought, high-salinity and cold stresses in order to survive. Molecular and genomic studies have shown that many genes with various functions are induced by drought, high-salinity and cold stresses, and that the various signaling factors including transcription factors are involved in the stress responses. The development of microarray-based expression profiling methods has allowed significant progress in the characterization of the plant stress response. Genetic engineering of the stress-inducible genes has become one of the promising strategies for the molecular breeding of the stress-tolerant plants. In this review, we highlight the application of the microarray analysis to the understanding of the plant abiotic stress responses and tolerance.

14.1 Introduction

Plant growth is greatly affected by environmental abiotic stresses, such as drought, high salinity and low temperature. Plants must adapt to these stresses in order to survive. Expression of a variety of genes is induced by these stresses in various plants and several hundred genes that respond to these stresses at the transcriptional level have been identified using microarray technology and so on (Shinozaki and Yamaguchi-Shinozaki 2000; Kreps et al. 2002; Seki et al. 2002a, 2003; Xiong et al.

M. Seki, M. Okamoto, A. Matsui, J.-M. Kim, Y. Kurihara, J. Ishida,
T. Morosawa, M. Kawashima, and T.K. To
Plant Genomic Network Research Team, Plant Functional Genomics Research Group, RIKEN Plant Science Center (PSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, 230-0045, Japan

M. Seki (✉)
Kihara Institute for Biological Research, Yokohama City University, 641-12 Maioka-cho, Totsuka-ku, Yokohama, 244-0813, Japan
e-mail: mseki@psc.riken.jp

K. Shinozaki
Gene Discovery Research Team, Gene Discovery Research Group, RIKEN Plant Science Center (PSC), RIKEN Tsukuba Institute, 3-1-1 Koyadai, Tsukuba, Ibaraki, 305-0074, Japan

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_14,

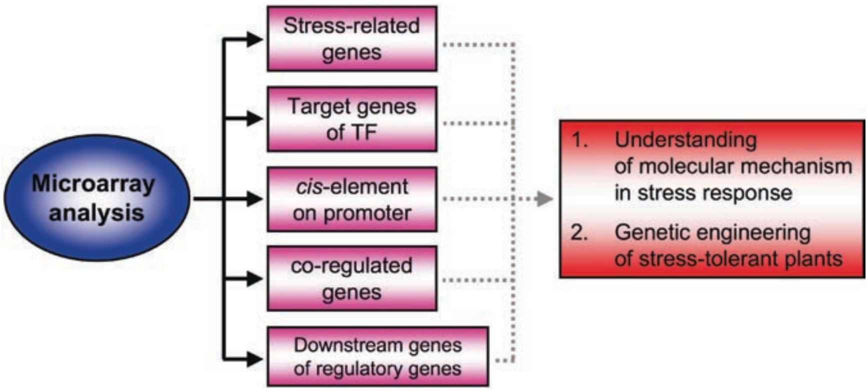


Fig. 14.1 The application of microarray analysis

2002; Zhu 2002; Shinozaki et al. 2003; Lee et al. 2005; Fig. 14.1). It is important to analyze the functions of stress-inducible genes not only to understand the molecular mechanisms of stress tolerance and the responses of higher plants but also to improve the stress tolerance of crops by gene manipulation. Stress-inducible genes have been used to improve the stress tolerance of transgenic plants (Zhang 2003; Umezawa et al. 2006; Valliyodan and Nguyen 2006; Fig. 14.1). Identification of the stress-co-regulated genes by microarray analysis is useful for identifying the *cis*-acting elements (Simpson et al. 2003) and helps us to understand the molecular mechanisms of the stress responses (Fig. 14.1).

DNA microarrays are useful for identifying the downstream target genes of the stress-related transcription factors and also for identifying potential *cis*-acting DNA-elements by combining the expression data with the genomic sequence data (Seki et al. 2001; Maruyama et al. 2004; Fig. 14.1). Many knock-out mutants with the stress-tolerant and the stress-sensitive phenotypes have been identified by genetic or reverse genetic approaches (Bartels and Sunkar 2005). Microarray analysis of the mutants have identified the genes that are regulated at the transcriptional level by the signaling components and helped us understand the regulatory networks in the abiotic stress responses (Fig. 14.1).

In this chapter, we summarize recent progress on the microarray analysis for the drought, cold and high-salinity stress responses in a model plant, *Arabidopsis*.

14.2 Identification of the Genes Upregulated by the Stresses

Microarray technology is a powerful tool for the systematic analysis of expression profiles of large numbers of genes (Richmond and Somerville 2000; Seki et al. 2004). Several hundred genes that respond to drought, cold and high-salinity stresses have been identified at the transcriptional level by microarray technology in *Arabidopsis* (Fowler and Thomashow 2002; Kreps et al. 2002; Seki et al. 2002a;

Lee et al. 2005). Abiotic stress-responsive genes have also been identified in other plant species, such as an *Arabidopsis*-related halophyte, *Thellungiella halophila* (Inan et al. 2004; Taji et al. 2004; Gong et al. 2005; Wong et al. 2006), barley (Oztur et al. 2002), chickpea (Mantri et al. 2007), grapevine (Cramer et al. 2007), hot pepper (Hwang et al. 2005), maize (Wang et al. 2003; Yu and Setter 2003), pine (Watkinson et al. 2003), *Poncirus trifoliata* (Sahin-Cevik and Moore 2006), poplar (Gu et al. 2004; Brosche et al. 2005), potato (Rensink et al. 2005), rice (Kawasaki et al. 2001; Rabbani et al. 2003; Lan et al. 2005), sorghum (Buchanan et al. 2005) and wheat (Gulick et al. 2005; Kawaura et al. 2006).

The products of the drought-, high-salinity- and cold-stress-inducible genes can be classified into two groups (Shinozaki and Yamaguchi-Shinozaki 2000; Seki et al. 2002a; 2003). The first group includes functional proteins that probably function in stress tolerance. They include late embryogenesis abundant (LEA) proteins, chaperones, water channel proteins, transporters, detoxification enzymes, key enzymes for osmolyte biosynthesis and proteases. The second group contains regulatory proteins, that is, protein factors involved in further regulation of signal transduction and gene expression that probably function in the stress response (Shinozaki and Yamaguchi-Shinozaki 2000; Seki et al. 2002a, 2003). They include various transcription factors, protein kinases, protein phosphatases, enzymes involved in phospholipid metabolism, F-box proteins, and other signaling molecules, such as calmodulin-binding protein (Seki et al. 2002a). Many stress-inducible transcription factor genes were identified, suggesting that various transcriptional regulatory mechanisms function in regulating drought, cold, or high salinity stress signal transduction pathways. Several reviews on the transcriptome in abiotic stress conditions in higher plants have been published recently (Bray 2002; Ramanjulu and Bartels 2002; Hazen et al. 2003; Seki et al. 2003, 2004, 2005; Yamaguchi-Shinozaki and Shinozaki 2005, 2006).

14.3 Transcriptome Analysis for the Stress-Inducible Transcription Factor Genes

14.3.1 Stress-Inducible Transcription Factors

Transcription factors play important roles in plant responses to environmental stresses and its development. Transcription factors are sequence-specific DNA-binding proteins that are capable of activating and/or repressing transcription. The *Arabidopsis* genome encodes more than 1,500 transcription factors (Riechmann et al. 2000; Iida et al. 2005) and a number of transcription factor families have been implicated in plant stress responses (Shinozaki et al. 2003; Yamaguchi-Shinozaki and Shinozaki 2005, 2006). Microarray analysis has also revealed many stress-inducible transcription factor genes and demonstrated overlap among various stress- or hormone-signaling pathways (Seki et al. 2001, 2002a, b, 2004; Chen et al.

2002; Cheong et al. 2002; Shinozaki et al. 2003). The major transcription factor families that are induced by the abiotic stresses are ERF/AP2, bZIP, NAC, MYB, bHLH, Cys2Cys2 zinc-finger, Cys2His2 zinc-finger, WRKY, HB and HSF. These transcription factors probably govern expression of stress-inducible genes either cooperatively or independently, and may constitute gene networks in drought, cold and high-salinity stress responses in *Arabidopsis*.

14.3.2 Identifying the Target Genes of Transcription Factors

DNA microarrays are useful for identifying the downstream target genes of the stress-related transcription factors. The stress-inducible transcription factors and *cis*-acting elements in the promoters of its target genes function in the signal transduction network from perception of stress signals to stress-responsive gene expression. DNA microarrays are also useful to identify potential *cis*-acting DNA elements by combining the expression data with the genomic sequence data in several plant species, such as *Arabidopsis* and rice in which the complete genomic sequence was determined.

Target genes of many abiotic stress-related transcription factors in *Arabidopsis* have been studied by microarray analysis (Table 14.1). The microarray analysis has also been used for identification of the target genes of rice (OsDREB1A, Dubouzet et al. 2003; Ito et al. 2006) and *Brassica* (DREB1/CBF homologs, BNCBF5 and 17, Savitch et al. 2005) stress-related transcription factors. Recent progress in transcriptome analysis for the stress-related transcription factors is summarized below.

Transcription factors belonging to the ERF/AP2 family that bind to the DRE/CRT elements were isolated and termed DREB1/CBF and DREB2 (Yamaguchi-Shinozaki and Shinozaki 2005). Their conserved DNA-binding motif is A/GCCGAC. The DREB1/CBF genes are rapidly and transiently induced by cold stress, the products of which activate the expression of the target stress-inducible genes (Jaglo-Ottosen et al. 1998; Liu et al. 1998; Kasuga et al. 1999). Overexpression of DREB1/CBF in transgenic plants increased stress tolerance to freezing, drought and salt stresses, indicating that the DREB1/CBF proteins function in the acquirement of the stress tolerance (Liu et al. 1998). More than 40 target genes of DREB1/CBF have been identified by both cDNA and GeneChip microarray analysis (Seki et al. 2001; Fowler and Thomashow 2002; Maruyama et al. 2004; Vogel et al. 2005). The downstream target genes include C2H2 zinc-finger-type- and ERF/AP2-type- TFs, RNA-binding proteins, sugar transport proteins, LEA proteins, KIN (cold-inducible) proteins, RFO (raffinose family oligosaccharides)-biosynthesis-related proteins and protease inhibitors. Conserved sequences in the promoter regions of the DREB1A/CBF3 target genes were searched, and A/GCCGACNT was found in their promoter regions between -51 and -450 as a consensus DRE (dehydration-responsive element) (Maruyama et al. 2004).

Constitutive expression of a zinc-finger-type transcription factor, *ZAT12/RHL41* in *Arabidopsis* caused a small, but reproducible, increase in freezing tolerance

Table 14.1 Microarray analysis of the Arabidopsis genes involved in the abiotic stress responses

Genes	AGI code	Coded protein	Mutant or Transgenic	Phenotype of mutants or transgenic plants	Microarrays used	Profiling results	References
Transcription factors (TFs)							
DREB1A/CBF3	At4g25480	ERF/AP2 TF	DREB1A/CBF3 OX line	Increased tolerance to drought and freezing stress	1.3K cDNA	Up of 12 genes	Seki et al. (2001)
			DREB1A/CBF3 OX line	Increased tolerance to drought and freezing stress	8K Affymetrix	Up of 41 genes down of 27 genes	Fowler and Thomashow (2002)
			DREB1A/CBF3 OX line	Increased tolerance to drought and freezing stress	7K cDNA 8K Affymetrix	Up of 38 genes	Maruyama et al. (2004)
DREB1B/CBF1	At4g25490	ERF/AP2 TF	DREB1B/CBF1 OX line	Increased tolerance to freezing stress	8K Affymetrix	Up of 41 genes down of 27 genes	Fowler and Thomashow (2002)
DREB1C/CBF2	At4g25470	ERF/AP2 TF	DREB1C/CBF2 OX line		8K Affymetrix	Up of 41 genes down of 27 genes	Fowler and Thomashow (2002)
			DREB1C/CBF2 OX line		24K Affymetrix	Up of 151 genes down of 43 genes	Vogel et al. (2005)
DREB2A	At5g05410	ERF/AP2 TF	DREB2A OX line (active form with internal deletion)	Increased tolerance to drought stress	7K cDNA	Up of 17 genes	Sakuma et al. (2006a)
			DREB2A OX line (active form with internal deletion)	Increased tolerance to heat stress	22K Agilent	Up of 483 genes	Sakuma et al. (2006b)

(continued)

Table 14.1 (continued)

Genes	AGI code	Coded protein	Mutant or Transgenic	Phenotype of mutants or transgenic plants	Microarrays used	Profiling results	References
			T-DNA KO line	Decreased tolerance to heat stress	22K Agilent	Down of 11 genes for drought down of 50 genes for heat	Sakuma et al. (2006)
AREB1/ ABF2	At1g45249	bZIP TF	AREB1/ABF2 OX line (active form with internal deletion)	Increased tolerance to drought stress, hypersensitive to ABA	22K Agilent	Up of 8 genes	Fujita et al. (2005)
			AREB1/ABF2 OX line (phosphorylated active form)	Increased tolerance to drought stress	22K Agilent	Up of 26 genes	Furihata et al. (2006)
AtbZIP60	At1g42990	bZIP TF	AtbZIP60 OX line	Increased tolerance to salt stress	44K Agilent	Up of 29 genes	Fujita et al. (2007)
AtMYB2	At2g47190	MYB TF	AtMYB2 and AtMYC2 OX line	Increased tolerance to osmotic stress, hypersensitive to ABA	7K cDNA	Up of 32 genes	Abe et al. (2003)
AtMYB60	At1g08810	MYB TF	T-DNA KO line	Increased tolerance to drought stress	7K cDNA	Up of 6 genes down of 30 genes	Cominelli et al. (2005)
HOS10	At1g35515	MYB TF	T-DNA KO line	Hypersensitive to freezing and salt stress	24K Affymetrix	Up of 6 genes down of 6 genes	Zhu et al. (2005)
AtMYB41	At4g28110	MYB TF	AtMYB41 OX line	Increased sensitivity to drought stress	24K Affymetrix	Up of 149 genes down of 28 genes	Cominelli et al. (2008)

AtMYC2/ JIN1	At1g32640	bHLH TF	AtMYB2 and AtMYC2 OX line	Increased tolerance to osmotic stress, hypersensitive to ABA	7K cDNA	Up of 32 genes	Abe et al. (2003)
ICE1	At3g26744	bHLH TF	EMS-mutagenized T-DNA KO line	Decreased cold acclimation	24K Affymetrix	Up or down of 369 genes	Chinnusamy et al. (2003), Lee et al. (2005)
ZAT12/ RHL41	At5g59820	C2H2-Type zinc finger TF	ZAT12/RHL41 OX line	Increased freezing tolerance	24K Affymetrix	Up of 47 genes down of 158 genes	Vogel et al. (2005)
			ZAT12/RHL41 OX line		24K Affymetrix	Up of 42 genes down of 6 genes	Davletova et al. (2005)
ZFHD1	At1g69600	Zinc finger HD TF	ZFHD1 OX line	Increased tolerance to drought stress	22K Agilent	Up of 29 genes	Tran et al. (2007a)
ANAC019	At1g52890	NAC TF	ANAC019 OX line	Increased tolerance to drought stress	7K cDNA	Up of 17 genes	Tran et al. (2004)
ANAC055	At3g15500	NAC TF	ANAC055 OX line	Increased tolerance to drought stress	7K cDNA	Up of 14 genes	Tran et al. (2004)
ANAC072/ RD26	At4g27410	NAC TF	RD26/ANAC072 OX line	Increased tolerance to drought stress	7K cDNA	Up of 28 genes	Tran et al. (2004)
			RD26/ANAC072 OX line	Hypersensitive to ABA	22K Agilent	Up of 20 genes	Fujita et al. (2004)
			RD26/ANAC072-repressed line	Insensitive to ABA	22K Agilent	Down of 15 genes	Fujita et al. (2004)
HOS9	At2g01500	HD TF	EMS-mutagenized T-DNA KO line	Sensitive to freezing stress	24K Affymetrix	Up of 140 genes	Zhu et al. (2004)

(continued)

Table 14.1 (continued)

Genes	AGI code	Coded protein	Mutant or Transgenic	Phenotype of mutants or transgenic plants	Microarrays used	Profiling results	References
Protein kinases (PKs)							
SRK2C/ SnRK2.8	At1g78290	SNF1-related PK 2	SRK2C OX line	Increased tolerance to drought stress	22K Agilent	Up of 18 genes down of 14 genes	Umezawa et al. (2004)
SOS2/ CIPK24/ SnRK3.11	At5g35410	SNF1-related PK 3	EMS-mutagenized mutant	Hypersenstive to salt stress	22K Agilent	Up of 27 genes down of 3 genes	Kamei et al. (2005)
AtMKK2	At4g29810	MAP kinase kinase 2	MKK2 OX line	Increased tolerance to freezing and salt stress	24K Affymetrix	Up of 127 genes down of 25 genes	Teige et al. (2004)
RPK1	At1g69270	LRR receptor-like PK	T-DNA KO line	Insensitive to ABA	22K Agilent	Down of 39 genes	Osakabe et al. (2005)
			Chimeric RPK1 LRR and BRI1 kinase domain OX line	Increased sensitivity to ABA	22K Agilent	Up of 27 genes down of 19 genes	Osakabe et al. (2005)
AHK1/ ATHK1	At2g17820	Histidine kinase	T-DNA and Ds KD line	Hypersensitive to drought stress, ABA-insensitive germination	22K Agilent	Down of 190 and 120 genes for 2.5 and 9 h drought treatment, respectively	Tran et al. (2007b)
AHK2	At5g35750	Histidine kinase	Double mutant of T-DNA KO lines, *ahk2* and *ahk3*	Increased tolerance to drought and salt stress	22K Agilent	Up of 40 genes	Tran et al. (2007b)
AHK3	At1g27320	Histidine kinase	Double mutant of T-DNA KO lines, *ahk2* and *ahk3*	Increased tolerance to drought and salt stress	22K Agilent	Up of 40 genes	Tran et al. (2007b)

Others							
ADA2b	At4g16420	Transcriptional adaptor protein	T-DNA KO line	Increased tolerance to freezing stress	8K Affymetrix	Up of 176 genes down of 81 genes	Vlachonasios et al. (2003)
MKP1	At3g55270	MAP kinase phosphatase	T-DNA KO line	Hypersensitive to genotoxic stress, increased tolerance to salt stress	8K Affymetrix	Up of 21 genes	Ulm et al. (2002)
SOS3/CBL4	At5g24270	Protein with the calcineurin B subunit	EMS-mutagenized plant	Hypersenstive to salt stress	22K Agilent	Up of 4 genes down of 1 gene	Kamei et al. (2005)
ADR1	At1g33560	Disease resistance protein (CC-NBS-LRR class)	ADR1 OX line	Increased tolerance to drought stress	7K cDNA	Up of 20 genes	Chini et al. 2004
SLH1	At5g45260	Disease resistance protein (TIR-NBS-LRR-WRKY class)	Ds KO line	Sensitive to low humidity	7K cDNA	Up of several genes	Noutoshi et al. (2005)
ESK1	At3g55990	DUF231 domain-containing protein	EMS-mutagenized plant and T-DNA KO line	Increased tolerance to freezing stress	24K Affymetrix	Up of 173 genes down of 139 genes	Xin et al. (2007)
ABH1/CBP80	At2g13540	mRNA cap binding protein	T-DNA KO line	ABA-hypersensitive stomatal closing and Reduced wilting during drought	8K Affymetrix	Up of 13 genes down of 18 genes	Hugouvieux et al. (2001)

(continued)

Table 14.1 (continued)

Genes	AGI code	Coded protein	Mutant or Transgenic	Phenotype of mutants or transgenic plants	Microarrays used	Profiling results	References
AHG2	At1g55870	Poly(A)-specific ribonuclease	EMS-mutagenized plant	ABA-hypersensitive germination	7K cDNA	Up of 19 genes	Nishimura et al. (2005)
AtPARP2	At2g31320	Poly(ADP-ribose) polymerase	Transgenic plants with dsRNA construct containing the 5′ end of PARP2 in the stem structure	Increased tolerance to drought, high light and Heat stress	24K Affymetrix	Up or down of 886 genes	Vanderauwera et al. (2007)

(Vogel et al. 2005). Microarray analysis showed that many cold-, oxidative- or light stress-responsive genes are upregulated in the overexpressor (Davletova et al. 2005; Vogel et al. 2005), indicating an important role for the ZAT12 regulon in cold acclimation and reactive oxygen stress signaling. ZAT12 also downregulated the expression of the *DREB1/CBF* genes (Vogel et al. 2005), indicating a role for ZAT12 in a negative regulatory circuit that dampens expression of the DREB1/CBF cold response pathway.

The *INDUCER OF CBF EXPRESSION1 (ICE1)* gene was identified through the map-based cloning of the *Arabidopsis ice1* mutation, which affected the expression of the *CBF3/DREB1A* promoter::LUC transgenic plants (Chinnusamy et al. 2003). ICE1 encodes a MYC-type bHLH transcription factor that regulates the expression of *CBF3/DREB1A* genes. Microarray analysis showed that among 939 cold-regulated genes, the expression of 369 genes including many transcription factor ones was affected in the *ice1* mutant (Lee et al. 2005), suggesting that ICE1 may control many cold-responsive regulons. Overexpression of *ICE1* in transgenic plants resulted in improved freezing tolerance, supporting an important role for ICE1 in the cold stress response (Chinnusamy et al. 2003).

The ERF/AP2-type transcription factor DREB2 also recognizes the DRE sequence and is thought to be a major transcription factor functioning under drought- and high-salinity-stress conditions (Liu et al. 1998; Sakuma et al. 2002). Overexpression of the constitutive active forms of *DREB2A* (35S:*DREB2A* CA) increased tolerance to drought (Sakuma et al. 2006a) and heat stress (Sakuma et al. 2006b), and induces expression of many stress-responsive genes (Sakuma et al. 2006a, b). The downstream target genes include ones encoding LEA proteins, dehydrins, COR15A, RFO (raffinose family oligosaccharides)-biosynthesis-related protein, and heat-shock-related proteins. These results indicate that DREB2A functions in both water and heat-stress responses.

The bZIP-type transcription factors ABRE-binding protein (AREB)/ABRE-binding factor (ABF) can bind to ABRE (ABA-responsive element; PyACGTGGC) and activate ABA-dependent gene expression (Uno et al. 2000; Choi et al. 2000). Overexpression of *ABF3* and *ABF4/AREB2* resulted in ABA-hypersensitive and drought stress tolerance phenotypes in *Arabidopsis* (Kang et al. 2002). Overexpression of *AREB1/ABF2* in transgenic plants is not necessary to activate its downstream genes such as *RD29B*. Transgenic *Arabidopsis* plants overexpressing an active *AREB1* form with internal deletion for overcoming the masked transactivation activity showed ABA hypersensitivity and enhanced drought tolerance, and eight genes in two groups were upregulated: LEA class genes and ABA- and dehydration-inducible regulatory genes such as linker histone H1 and AAA ATPase (Fujita et al. 2005). These eight upregulated genes carry two or more ABRE motifs in the promoter regions. A Myc transcription factor, AtMYC2 (JIN1, Jasmonate Insensitive 1), and a MYB transcription factor, AtMYB2, bind to *cis*-elements in the promoter of the drought-inducible gene *RD22*, and activate *RD22* in a cooperative manner (Abe et al. 1997). The transgenic plants overexpressing *AtMYC2* and/or *AtMYB2* cDNAs have higher sensitivity to ABA (Abe

et al. 2003). Microarray analysis of the transgenic plants overexpressing *AtMYC2* and/or *AtMYB2* revealed that several ABA-inducible genes carrying the MYC recognition sequence (CANNTG) and the MYB recognition sequences (A/TAACCA and C/TAACG/TG) in the promoter regions were upregulated in the transgenic plants. *Ds* insertion mutant of the *AtMYC2* gene was less sensitive to ABA and showed significantly decreased ABA-induced gene expression of *RD22* and *AtADH1*. These results indicated that both AtMYC2 and AtMYB2 function as transcriptional activators in ABA-inducible gene expression under drought stress conditions in plants.

A MYB transcription factor, *AtMYB60*, is specifically expressed in guard cells, and its expression is negatively modulated during drought (Cominelli et al. 2005). A null mutation in *AtMYB60* results in the constitutive reduction of stomatal opening and in decreased wilting under water stress conditions. A microarray experiment showed that a number of genes such as the aquaporin delta-tonoplast intrinsic protein gene (δ-*TIP*) is upregulated in the mutant, indicating that *AtMYB60* is a transcriptional modulator of physiological responses in guard cells (Cominelli et al. 2005).

A mutation in a MYB transcription factor, *HOS10*, results in high-level expression of *RD29A* and other stress-responsive genes by ABA, low temperature and salt stress (Zhu et al. 2005). The *hos10* plants are extremely sensitive to freezing temperatures, completely unable to acclimate to the cold, and are hypersensitive to NaCl. The transcript level of *NCED3* encoding a key enzyme in ABA biosynthesis is much less in *hos10* than in wild-type after PEG-induced dehydration stress, suggesting that HOS10 controls expression of genes involved in ABA biosynthesis under dehydration stress (Zhu et al. 2005).

The transgenic plants overexpressing a MYB transcription factor, *AtMYB41*, showed a dwarf appearance, dependent on smaller cells with abnormal morphology, enhanced sensitivity to desiccation, and enhanced permeability of leaf surfaces, suggesting discontinuity in the cuticle (Cominelli et al. 2008). *AtMYB41*-regulated genes included ones involved in lipid metabolism and transport, cell-wall modifications, and cuticle metabolism, suggesting that AtMYB41 partly controls cell expansion and cuticle deposition in response to abiotic stress.

A mutation in a homeodomain transcription factor, *HOS9*, results in high-level expression of *RD29A* by low temperature, but not by ABA or salinity stress (Zhu et al. 2004). The *hos9* plants are more sensitive to freezing than the wild-type plants. Microarray analysis showed that none of the genes affected by *hos9* mutation are controlled by the DREB1/CBF family (Zhu et al. 2004), suggesting that HOS9 is important for freezing tolerance by affecting the activity of genes independent of the DREB1/CBF pathway.

Several dehydration-inducible genes, such as early response to dehydration 1 (*ERD1*) encoding a Clp protease regulatory subunit, ClpD, do not respond to either cold or ABA treatment. Promoter analysis of *ERD1* revealed that two different novel *cis*-acting elements, a MYC-like sequence (CATGTG) and a 14-bp *rps1* site 1-like sequence (CACTAAATTGTCAC), are involved in induction by dehydration stress (Simpson et al. 2003). Three cDNAs encoding the MYC-like

sequence-binding proteins, *ANAC019*, *ANAC055* and *ANAC072*, which belong to the NAC transcription factor family, were isolated by the yeast one-hybrid screening method (Tran et al. 2004). Microarray analysis of the transgenic plants overexpressing either *ANAC019*, *ANAC055*, or *ANAC072/RD26* revealed that several stress-inducible genes were upregulated in the transgenic plants (Fujita et al. 2004; Tran et al. 2004), and the plants showed significantly increased drought tolerance (Tran et al. 2004). However, *ERD1* was not upregulated in the transgenic plants. A cDNA encoding the 14-bp-*rps1*-site 1-like sequence-binding protein, ZFHD1, which belongs to the zinc-finger homeodomain (ZFHD) transcription factor family, was isolated by the yeast one-hybrid screening method (Tran et al. 2007a). *ZFHD1*-overexpressing plants displayed an increased tolerance to drought stress. Expression of *ERD1* was not upregulated in the overexpressors of the *ZFHD1*, but upregulated in the co-overexpressors of both the *ZFHD1* and *NAC* genes. These results indicate that both *cis*-acting elements are necessary for expression of *ERD1*.

Several recent reviews are available on the roles of the transcription factors in the abiotic stress signaling and the expression profiling results (Shinozaki et al. 2003; Bartels and Sunkar 2005; Seki et al. 2005; Yamaguchi-Shinozaki and Shinozaki 2005, 2006; Umezawa et al. 2006). Information on the target genes is useful for understanding the transcriptional regulatory networks in cellular responses to the abiotic stresses.

14.4 Analysis of the Transcriptome Regulated by the Regulatory Proteins

Molecular and genetic studies have shown that various signal transduction systems function in abiotic stress responses, involving protein phosphorylation and/or dephosphorylation, phospholipid signaling, calcium signaling, protein degradation and so on (Bartels and Sunkar 2005; Boudsocq and Lauriere 2005; Mahajan and Tuteja 2005; Vinocur and Altman 2005). Although these complex signaling processes are not yet fully understood, several genes encoding the signaling factors involved in the abiotic stress responses have been identified (Shinozaki et al. 2003; Chinnusamy et al. 2004; Bartels and Sunkar 2005; Umezawa et al. 2006). Recent progress in the microarray analysis for the stress-related signaling factors is summarized below (Table 14.1).

The *eskimo1* (*esk1*) mutation resulted in a constitutive freezing-tolerant phenotype in the absence of cold acclimation (Xin and Browse 1998). ESK1 encodes a DUF231-domain-containing proteins of unknown function (Xin et al. 2007). Microarray analysis revealed that 312 ESK1-regulated genes show greater overlap with the genes regulated by salt, osmotic and ABA treatments than with ones regulated by cold acclimation or by the DREB1A/CBF3. These results indicate that ESK1 encodes a novel negative regulator of cold acclimation and that the *esk1*

mutations mediate freezing tolerance through mechanisms that are substantially distinct from those controlled by the DREB1/CBF regulons.

Gain- and loss-of-function studies in *Arabidopsis* indicated that a histidine kinase, AHK1/ATHK1, is a positive regulator of drought and salt stress responses and ABA signaling (Urao et al. 1999; Tran et al. 2007b). Microarray analysis of the *athk1* mutant revealed a down regulation of many stress- and/or ABA-inducible genes, including *AREB1*, *ANAC* and *DREB2A* transcription factors and their downstream genes, suggesting that AHK1 functions upstream of *AREB1*, *ANAC* and *DREB2A*. The loss-of-function mutants of the cytokinin receptor histidine kinases AHK2 and AHK3, *ahk2*, *ahk3* and *ahk2 ahk3* were strongly tolerant to drought and salt stress due to upregulation of many stress- and/or ABA-inducible genes, suggesting that AHK2 and AHK3 negatively controls osmotic stress responses (Tran et al. 2007b). In the presence of cytokinin, the *cre1* mutant displayed a strong salt stress-tolerant phenotype, although the *cre1* mutant did not show phenotypic changes under drought and salt stress, suggesting that CRE1 also function in the stress responses as a negative regulator. Cross-talk may exist among cytokinin, ABA and stress signaling pathways.

Transcriptional activation by CBF1/DREB1B might be mediated, at least in part, by homologs of the yeast ADA2 and GCN5 proteins that encode components of the ADA and SAGA histone acetyltransferase complexes (Stockinger et al. 2001). Nonacclimated *ada2b* mutants were more freezing tolerant than nonacclimated wild-type plants (Vlachonasios et al. 2003). In cold acclimation experiments, *DREB1/CBF* genes were induced in the *ada2b* mutants as in wild-type plants, but subsequent transcription of cold-regulated (*COR*) genes was reduced in the mutants.

A SNF1-related protein kinase 2 (SnRK2), SRK2C/SnRK2.8, is an osmotic-stress-activated protein kinase (Umezawa et al. 2004). Knock-out mutants of *SRK2C* exhibited drought hypersensitivity in the roots and the overexpressors of *SRK2C* displayed higher overall drought tolerance than control plants. Microarray analysis of the overexpressors revealed that their drought tolerance coincided with upregulation of many stress-responsive genes, such as *RD29A*, *COR15A* and *DREB1A/CBF3*. These studies demonstrate that activation of SRK2C mediates drought-stress signaling and improves the drought tolerance of *Arabidopsis* plants.

Under salt-stress conditions, *Arabidopsis salt overly sensitive* (*sos*) mutants show hypersensitive root growth (Liu and Zhu 1998; Liu et al. 2000). Microarray analysis of the *sos2* and *sos3* mutants revealed that the expression level of the major salt-inducible genes (*RD29A/COR78*, *RD17/COR47*, *COR15A*, *KIN1* and *RD22*) in the *sos* mutants were similar to those in the wild-type plants, indicating that the SOS2/SOS3 signaling pathway is independent of the DRE/CRT, ABRE and MYC/MYB pathways (Kamei et al. 2005).

The plants overexpressing an *Arabidopsis* mitogen-activated protein kinase (MAPK) kinase 2, MKK2, exhibited constitutive MPK4 and MPK6 activity, constitutively upregulated expression of the stress-induced marker genes, and increased freezing and salt tolerance (Teige et al. 2004). Microarray analysis of the *MKK2*-

overexpressing plants demonstrated upregulated expression of 127 genes involved in transcriptional regulation, signal transduction, cellular defense and stress metabolism. These results indicate that MKK2 plays a critical role in the cold and salt stress responses in *Arabidopsis*.

A mutation in a MAP kinase phosphatase, *MKP1*, results in the phenotypes of salt-tolerance and hypersensitivity to genotoxic stress (Ulm et al. 2002). Microarray analysis of the mutants revealed upregulated expression of the abiotic-, biotic- or oxidative-stress-related genes, indicating that MKP1 plays a pivotal role in the integration and fine-tuning of plant responses to various environmental challenges.

A Leu-rich repeat (LRR) receptor-like kinase in the plasma membrane, *RPK1*, is upregulated by ABA in *Arabidopsis*. Repression of *RPK1* expression decreased sensitivity to ABA during germination, growth, and stomatal closure (Osakabe et al. 2005). Microarray analysis showed that many ABA-inducible genes are downregulated in these plants. These results indicate an important role of RPK1 in ABA signal transduction.

Constitutive or conditional enhanced expression of the gene encoding a coiled-coil (CC)-nucleotide-binding site (NBS)-leucine-rich repeat (LRR) protein, ADR1, conferred drought and broad spectrum disease resistance (Chini et al. 2004). Microarray analysis of the plants containing a conditional *adr1* allele showed that salicylic acid-dependent genes, such as *PR1* and *PR5*, and the antioxidant responses-related genes, such as glutathione S-transferase are upregulated. These results indicate that overlap exists between signaling networks establishing disease resistance and drought tolerance.

Hugouvieux et al. (2001) isolated a recessive ABA-hypersensitive *Arabidopsis* mutant, *abh1*. ABH1 encodes a functional mRNA cap binding protein. DNA chip experiments showed that 18 genes including *RD20, KIN2* and *COR15b* had significant and threefold reduced transcript levels in the *abh1* mutant, and seven of these genes are ABA-regulated in the wild type. Consistent with these results, *abh1* plants showed ABA-hypersensitive stomatal closing and reduced wilting during drought. Hugouvieux et al. (2001) showed ABA-hypersensitive cytosolic calcium increases in *abh1* guard cells. These results indicate a functional link between mRNA processing and modulation of early ABA signal transduction. Recent studies revealed that knock-out mutants of several components functioning in RNA processing, such as *hyl1* (Lu and Fedoroff 2000), *cbp20* (Papp et al. 2004), *ahg2* (Nishimura et al. 2005) and *sad1* (Xiong et al. 2001), exhibit ABA hypersensitivity. Hyl1, CBP20, AHG2 and SAD1 encode nuclear-localized double-stranded RNA binding protein, mRNA cap binding protein, poly(A)-specific ribonuclease and Sm-like snRNP protein, respectively.

The information on the transcriptome helps us understand the regulatory network of the abiotic stress responses: specificity and cross-talk. The roles of the signaling molecules in the abiotic stress signaling and the expression profiling results are summarized in recent reviews (Bartels and Sunkar 2005; Boudsocq and Lauriere 2005; Umezawa et al. 2006).

14.5 Genetic Engineering of Abiotic Stress Tolerance Using the Stress-Inducible Genes

Genetic engineering of the stress-inducible genes that have important roles in the stress responses is one of the most promising strategies for improvement of the stress tolerance in plants. Several reviews on the engineering the stress tolerance in plants have been published (Zhang 2003; Umezawa et al. 2006; Valliyodan and Nguyen 2006; Christensen and Feldmann 2007).

Several stress-inducible genes encoding the functional proteins have been used for the improvement of the stress tolerance (Umezawa et al. 2006; Valliyodan and Nguyen 2006). Improvement of the stress tolerance using the stress-inducible genes encoding the proteins involved in the metabolism of the osmoprotectants, such as galactinol (Taji et al. 2002), proline (Yamada et al. 2005) and polyamine (Kasukabe et al. 2004), the protective proteins, such as LEA proteins (Chandra Babu et al. 2004), the reactive oxygen species (ROS)-scavenging proteins (De Block et al. 2005), transporters (Shi et al. 2003) and ABA-metabolism-related proteins, such as a key ABA biosynthesis enzyme (NCED3) (Iuchi et al. 2001) have been reported.

Several transcriptional activators such as DREB1/CBF (Jaglo-Ottosen et al. 1998; Liu et al. 1998; Kasuga et al. 1999) that upregulate the stress-responsive genes have been utilized to produce the stress-tolerant transgenic plants (Zhang 2003; Umezawa et al. 2006). The *DREB1/CBF* genes have been successfully used to engineer abiotic stress tolerance in various plant species, such as canola (Jaglo-Ottosen et al. 2001), chrysanthemum (Hong et al. 2006), rice (Ito et al. 2006), peanut (Bhatnagar-Mathur et al. 2007) and potato (Behnam et al. 2007).

14.6 Conclusions and Future Perspectives

The microarray-based expression profiling method is useful for identifying the stress-inducible genes. Functional analysis of the stress-inducible genes has provided more information on the signal transduction in the stress responses. Genetic engineering of the stress-inducible genes has become one of the most promising strategies for improvement of the stress tolerance in plants. Expression profiling studies also provide the expression levels of many genes as a detailed snapshot that describes the state of a biological system in the plants under some conditions. The phenotypic descriptions of the mutants and the transgenic plants for the profiling of the stress-related genes can be used to infer the relationships among the genes and to understand the molecular mechanism in the stress responses.

Large *Arabidopsis* microarray data sets are available from GEO (http://www.ncbi.nlm.nih.gov/geo) (Edgar et al. 2002) and ArrayExpress (http://www.ebi.ac.uk/arrayexpress) (Brazma et al. 2003). User-friendly web-based *Arabidopsis* microarray data analysis tools, such as Genevestigator (http://www.genevestigator.ethz.ch/) (Zimmermann et al. 2004) to provide categorized quantitative information

about genes, KaPPA-View (http://kpv.kazusa.or.jp/kappa-view/) (Tokimatsu et al. 2005) to integrate the transcript and metabolite data, and ATTED-II (http://www.atted.bio.titech.ac.jp) (Obayashi et al. 2006) to identify the co-regulated gene groups are also available. Availability of large sets of the plant microarray data and user-friendly analytical tools should aid the functional analysis of the stress-related genes and our better understanding the transcriptional regulatory networks in the abiotic stress responses.

Recently, non-coding RNAs, such as small RNAs (Sunkar et al. 2007), alternative splicing (Iida et al. 2004; Lee et al. 2006) and chromatin remodeling (Sridha and Wu 2006; Kim et al. 2008) have been shown to be involved in the abiotic stress responses and tolerance (Seki et al. 2007). The whole genome tiling array (Yamada et al. 2003; Stolc et al. 2005; Matsui et al. 2008) is a powerful tool for identification of stress-regulated non-coding RNAs, and for analysis of alternative splicing and chromatin remodeling. New sequencing technologies developed by 454 Life Sciences (http://www.454.com/) (Margulies et al. 2005), Solexa, Inc. (http://www.solexa.com/) and Applied Biosystems, Inc. (http://www.appliedbiosystems.com/) will enable the identification of large numbers of the small RNAs involved in the abiotic stress responses. Comparative genomics studies using bioinformatic approaches will also lead to our better understanding of the function and biological significance of the non-coding RNAs such as small RNAs.

Recently, metabolome analysis has become a useful approach to study metabolic changes in response to stress (Cook et al. 2004; Rizhsky et al. 2004; Urano et al. unpublished results). Metabolome profiling experiments enable us to identify a large number of metabolites that increased or decreased during the stress conditions. Integration of metabolome data with the transcriptome data during the stress condition should help us identify the key metabolic pathway genes for genetic engineering of the stress tolerance and find novel pathways in the stress response.

Acknowledgements This work was supported in part by a grant for Genome Research from RIKEN, the Program for Promotion of Basic Research Activities for Innovative Biosciences, the Special Coordination Fund of the Science and Technology Agency, and a Grant-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MECSST) to K.S. It was also supported in part by a Grant-in-Aid for Scientific Research on Priority Areas "Systems Genomics" from MECSST and the President Discretionary Fund from RIKEN to M.S

References

Abe H, Yamaguchi-Shinozaki K, Urao T et al. (1997) Role of *Arabidopsis* MYC and MYB homologs in drought- and abscisic-acid-regulated gene expression. Plant Cell 9:1859–1868

Abe H, Urao T, Ito T et al. (2003) *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell 15:63–78

Bartels D, Sunkar R (2005) Drought and salt tolerance in plants. Crit Rev Plant Sci 24:23–58

Behnam B, Kikuchi A, Celebi-Toprak F et al. (2007) Arabidopsis rd29A::DREB2A enhances freezing tolerance in transgenic potato. Plant Cell Rep 26:1275–1282

Bhatnagar-Mathur P, Devi MJ, Reddy DS et al. (2007) Stress-inducible expression of *AtDREB1A* in transgenic peanut (*Arachis hypogaea* L.) increases transpiration efficiency under water-limiting conditions. Plant Cell Rep 26:2071–2082

Boudsocq M, Lauriere C (2005) Osmotic signaling in plants. Multiple pathways mediated by emerging kinase families. Plant Physiol 138:1185–1194

Bray EA (2002) Classification of genes differentially expressed during water-deficit stress in *Arabidopsis thaliana*: an analysis using microarray and differential expression data. Ann Bot 89:803–811

Brazma A, Parkinson H, Sarkans U et al. (2003) ArrayExpress-a public repository for microarray gene expression data at the EBI. Nucleic Acids Res 31:68–71

Brosche M, Vinocur B, Alatalo ER et al. (2005) Gene expression and metabolite profiling of *Populus euphratica* growing in the Negev desert. Genome Biol 6:R101

Buchanan CD, Lim S, Salzman RA et al. (2005) *Sorghum bicolor*'s transcriptome response to dehydration, high salinity and ABA. Plant Mol Biol 58:699–720

Chandra Babu R, Zhang J, Blumc A et al. (2004) *HVA1*, a LEA gene from barley confers dehydration tolerance in transgenic rice (*Oryza sativa* L.) via cell membrane protection. Plant Sci 166:855–862

Chen W, Provart NJ, Glazebrook J et al. (2002) Expression profile matrix of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses. Plant Cell 14:559–574

Cheong YH, Chang HS, Gupta R et al. (2002) Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in *Arabidopsis*. Plant Physiol 129:661–677

Chini A, Grant J, Seki M et al. (2004) Drought tolerance established by enhanced expression of the CC-NBS-LRR gene, ADR1, requires salicylic acid, EDS1 and ABI1. Plant J 38:810–822

Chinnusamy V, Ohta M, Kannar S et al. (2003) ICE1: a regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. Genes Dev 17:1043–1054

Chinnusamy V, Schumaker K, Zhu JK (2004) Molecular genetic perspectives on cross-talk and specificity in abiotic stress signaling in plants. J Exp Bot 55:225–236

Choi H, Hong JH, Ha J et al. (2000) ABFs, a family of ABA-responsive element binding factors. J Biol Chem 275:1723–1730

Christensen CA, Feldmann KA (2007) Biotechnology approaches to engineering drought tolerant crops. In: Jenks MA, Hasegawa PM, Jain SM (eds) Advances in molecular breeding toward drought and salt tolerant crops. Springer, Dordrecht, pp 333–357

Cominelli E, Galbiati M, Vavasseur A et al. (2005) A guard-cell-specific MYB transcription factor regulates stomatal movements and plant drought tolerance. Curr Biol 15:1196–1200

Cominelli E, Sala T, Calvi D et al. (2008) Over-expression of the Arabidopsis *AtMYB41* gene alters cell expansion and leaf surface permeability. Plant J 15:1196–1200

Cook D, Fowler S, Fiehn O et al. (2004) A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of *Arabidopsis*. Proc Natl Acad Sci U S A 101: 15243–15248

Cramer GR, Ergul A, Grimplet J et al. (2007) Water and salinity stress in grapevines: early and late changes in transcript and metabolite profiles. Funct Integr Genomics 7:111–134

Davletova S, Schlauch K, Coutu J et al. (2005) The Zinc-finger protein Zat12 plays a central role in reactive oxygen and abiotic stress signaling in *Arabidopsis*. Plant Physiol 139:847–856

De Block M, Verduyn C, Brouwer DD et al. (2005) Poly (ADP-ribose) polymerase in plants affects energy homeostasis, cell death and stress tolerance. Plant J 41:95–106

Dubouzet JG, Sakuma Y, Ito Y et al. (2003) OsDREB genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. Plant J 33:751–763

Edgar R, Domrachev M, Lash AE (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30:207–210

Fowler S, Thomashow MF (2002) *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. Plant Cell 14:1675–1690

Fujita M, Fujita Y, Maruyama K et al. (2004) A dehydration-induced NAC protein, RD26 is involved in ABA-dependent stress signaling pathway. Plant J 39:863–876

Fujita Y, Fujita M, Sato R et al. (2005) AREB1 is a transcription activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in *Arabidopsis*. Plant Cell 17:3470–3488

Fujita M, Mizukado S, Fujita Y et al. (2007) Identification of stress-tolerance-related transcription-factor genes via mini-scale full-length cDNA over-expressor (FOX) gene hunting system. Biochem Biophys Res Commun 364:250–257

Furihata T, Maruyama K, Fujita Y et al. (2006) Abscisic acid-dependent multisite phosphorylation regulates the activity of a transcription activator AREB1. Proc Natl Acad Sci USA 103:1988–1993

Gong Q, Li P, Ma S et al. (2005) Salinity stress adaptation competence in the extremophile *Thellungiella halophila* in comparison with its relative *Arabidopsis thaliana*. Plant J 44:826–839

Gu R, Fonseca S, Puskas LG et al. (2004) Transcript identification and profiling during salt stress and recovery of *Populus euphratica*. Tree Physiol 24:265–276

Gulick PJ, Drouin S, Yu Z et al. (2005) Transcriptome comparison of winter and spring wheat responding to low temperature. Genome 48:913–923

Hazen SP, Wu Y, Kreps JA (2003) Gene expression profiling of plant responses to abiotic stress. Funct Integr Genomics 3:105–111

Hong B, Tong Z, Ma N et al. (2006) Heterologous expression of the *AtDREB1A* gene in chrysanthemum increases drought and salt stress tolerance. Sci China C Life Sci 49:436–445

Hugouvieux V, Kwak JM, Schroeder JI (2001) An mRNA cap binding protein, ABH1, modulates early abscisic acid signal transduction in *Arabidopsis*. Cell 106:477–487

Hwang EW, Kim KA, Park SC et al. (2005) Expression profiles of hot pepper (*Capsicum annuum*) genes under cold stress conditions. J Biosci 30:657–667

Iida K, Seki M, Sakurai T et al. (2004) Genome-wide analysis of alternative pre-mRNA splicing in *Arabidopsis thaliana* based on full-length cDNA sequences. Nucleic Acids Res 32:5096–5103

Iida K, Seki M, Sakurai T et al. (2005) RARTF: database and tools for complete sets of *Arabidopsis* transcription factors. DNA Res 12:247–256

Inan G, Zhang Q, Li P et al. (2004) Salt Cress. A halophyte and cryophyte Arabidopsis relative model system and its applicability to molecular genetic analyses of growth and development of extremophiles. Plant Physiol 135:1718–1737

Ito Y, Katsura K, Maruyama K et al. (2006) Functional analysis of rice DREB1/CBF-type transcription factors Involved in cold-responsive gene expression in transgenic rice. Plant Cell Physiol 47:141–153

Iuchi S, Kobayashi M, Taji T et al. (2001) Regulation of drought tolerance by gene manipulation of 9-*cis*-epoxycarotenoid dioxygenase, a key enzyme in abscisic acid biosynthesis, in *Arabidopsis*. Plant J 27:325–333

Jaglo-Ottosen KR, Gilmour SJ, Zarka DG et al. (1998) *Arabidopsis* CBF1 overexpression induces *cor* genes and enhances freezing tolerance. Science 280:104–106

Jaglo-Ottosen KR, Kleff S, Amundsen KL et al. (2001) Components of the *Arabidopsis* C-repeat/dehydration-responsive element binding factor cold-response pathway are conserved in *Brassica napus* and other plant species. Plant Physiol 127:910–917

Kamei A, Seki M, Umezawa T et al. (2005) Analysis of gene expression profiles in *Arabidopsis* salt overly sensitive mutants, *sos2* and *sos3* mutants. Plant Cell Environ 28:1267–1275

Kang JY, Choi HI, Im MY et al. (2002) Arabidopsis basic leucine zipper proteins that mediate stress-responsive abscisic acid signaling. Plant Cell 14:343–357

Kasuga M, Liu Q, Miura S et al. (1999) Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. Nat Biotechnol 17:287–291

Kasukabe Y, He L, Nada K et al. (2004) Overexpression of spermidine synthase enhances tolerance to multiple environmental stresses and upregulates the expression of various stress-regulated genes in transgenic *Arabidopsis thaliana*. Plant Cell Physiol 45:712–722

Kawasaki S, Borchert C, Deyholos M et al. (2001) Gene expression profiles during the initial phase of salt stress in rice. Plant Cell 13:889–905

Kawaura K, Mochida K, Yamazaki Y et al. (2006) Transcriptome analysis of salinity stress responses in common wheat using a 22k oligo-DNA microarray. Funct Integr Genomics 6:132–142

Kim JM, Kim-To T, Ishida J et al. (2008) Alterations of lysine modifications on histone H3 N-tail under drought stress conditions in *Arabidopsis thaliana*. Plant Cell Physiol 49:1580–1588

Kreps JA, Wu Y, Chang HS et al. (2002) Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress. Plant Physiol 130:2129–2141

Lan L, Li M, Lai Y et al. (2005) Microarray analysis reveals similarities and variations in genetic programs controlling pollination/fertilization and stress responses in rice (*Oryza sativa* L.). Plant Mol Biol 59:151–164

Lee BH, Henderson DA, Zhu JK (2005) The Arabidopsis cold-responsive transcriptome and its regulation by ICE1. Plant Cell 17:3155–3175

Lee BH, Kapoor A, Zhu J et al. (2006) STABILIZED1, a stress-upregulated nuclear protein, is required for pre-mRNA splicing, mRNA turnover, and stress tolerance in Arabidopsis. Plant Cell 18:1736–1749

Liu Q, Kasuga M, Sakuma Y et al. (1998) The transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in Arabidopsis. Plant Cell 10:1391–1406

Liu J, Zhu JK (1998) A calcium sensor homolog required for plant salt tolerance. Science 280:1943–1945

Liu J, Ishitani M, Halfter U et al. (2000) The Arabidopsis thaliana SOS2 gene encodes a protein kinase that is required for salt tolerance. Proc Natl Acad Sci U S A 97:3730–3734

Lu C, Fedoroff N (2000) A mutation in the *Arabidopsis* HYL1 gene encoding a dsRNA-binding protein affects responses to abscisic acid, auxin, and cytokinin. Plant Cell 12:2351–2366

Mahajan S, Tuteja N (2005) Cold, salinity and drought stresses: an overview. Arch Biochem Biophys 444:139–158

Mantri NL, Ford R, Coram TE et al. (2007) Transcriptional profiling of chickpea genes differentially regulated in response to high-salinity, cold and drought. BMC Genomics 8:303

Margulies M, Egholm M, Altman WE et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380

Maruyama K, Sakuma Y, Kasuga M et al. (2004) Identification of cold-inducible downstream genes of the Arabidopsis DREB1A/CBF3 transcriptional factor using two microarray systems. Plant J 38:982–993

Matsui A, Ishida J, Morosawa T et al. (2008) Arabidopsis transcriptome analysis under drought, cold, high-salinity and ABA treatment conditions using a tiling array. Plant Cell Physiol 49:1135–1149

Nishimura N, Kitahata N, Seki M et al. (2005) Analysis of *ABA Hypersensitive Germination2* revealed the pivotal functions of PARN in stress response in *Arabidopsis*. Plant J 44:972–984

Noutoshi Y, Ito T, Seki M et al. (2005) A single amino acid insertion in the WRKY domain of the *Arabidopsis* TIR-NBS-LRR-WRKY type disease resistance protein SLH1 (SENSITIVE TO LOW HUMIDITY 1) causes activation of defense responses and hypersensitive cell death. Plant J 43:873–888

Obayashi T, Kinoshita K, Nakai K et al. (2006) ATTED-II: a database of co-expressed genes and *cis* elements for identifying co-regulated gene groups in *Arabidopsis*. Nucleic Acids Res 35:D863–D869

Osakabe Y, Maruyama K, Seki M et al. (2005) An LRR receptor kinase, RPK1, is a key membrane-bound regulator of abscisic acid early signaling in *Arabidopsis*. Plant Cell 17:1105–1119

Oztur ZN, Talame V, Deyholos M et al. (2002) Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. Plant Mol Biol 48:551–573

Papp I, Mur LA, Dalmadi A et al. (2004) A mutation in the Cap Binding Protein 20 gene confers drought tolerance to *Arabidopsis*. Plant Mol Biol 55:679–686

Rabbani MA, Maruyama K, Abe H et al. (2003) Monitoring expression profiles of rice (*Oryza sativa* L.) genes under cold, drought and high-salinity stresses, and ABA application using both cDNA microarray and RNA gel blot analyses. Plant Physiol 133:1755–1767
Ramanjulu S, Bartels D (2002) Drought- and desiccation-induced modulation of gene expression in plants. Plant Cell Environ 25:141–151
Rensink WA, Lobst S, Hart A et al. (2005) Gene expression profiling of potato responses to cold, heat, and salt stress. Funct Integr Genomics 5:201–207
Richmond T, Somerville S (2000) Chasing the dream: plant EST microarrays. Curr Opin Plant Biol 3:108–116
Riechmann JL, Heard J, Martin G et al. (2000) *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. Science 290:2105–2110
Rizhsky L, Liang H, Shuman J et al. (2004) When defense pathways collide. The response of *Arabidopsis* to a combination of drought and heat stress. Proc Natl Acad Sci U S A 134:1683–1696
Sahin-Cevik M, Moore GA (2006) Identification and expression analysis of cold-regulated genes from the cold-hardy *Citrus* relative *Poncirus trifoliate* (L.) Raf. Plant Mol Biol 62:83–97
Sakuma Y, Liu Q, Dubouzet JG et al. (2002) DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. Biochem Biophys Res Commun 290:998–1009
Sakuma Y, Maruyama K, Osakabe Y et al. (2006a) Functional analysis of an *Arabidopsis* transcription factor, DREB2A, involved in drought-responsive gene expression. Plant Cell 18:1292–1309
Sakuma Y, Maruyama K, Qin F et al. (2006b) Dual function of an Arabidopsis transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression. Proc Natl Acad Sci U S A 103:18822–18827
Savitch LV, Allard G, Seki M et al. (2005) The effects of overexpression of two *Brassica* CBF/DREB1-like transcription factors on phytosynthetic capacity and freezing tolerance in *Brassica napus*. Plant Cell Physiol 46:1525–1539
Seki M, Narusaka M, Abe H et al. (2001) Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses using a full-length cDNA microarray. Plant Cell 13:61–72
Seki M, Narusaka M, Ishida J et al. (2002a) Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold, and high-salinity stresses using a full-length cDNA microarray. Plant J 31:279–292
Seki M, Ishida J, Narusaka M et al. (2002b) Monitoring the expression pattern of ca. 7000 Arabidopsis genes under ABA treatments using a full-length cDNA microarray. Funct Integr Genomics 2:282–291
Seki M, Kamei A, Satou M et al. (2003) Transcriptomeanalysis in abiotic stress conditions in higher plants. Topics Curr Genet 4:271–295
Seki M, Satou M, Sakurai T et al. (2004) RIKEN Arabidopsis full-length (RAFL) cDNA and its applications for expression profiling under abiotic stress conditions. J Exp Bot 55:213–223
Seki M, Ishida J, Nakajima M et al. (2005) Genomic analysis of stress response. In: Jenks MA, Hasegawa PM (eds) Plant abiotic stress. Blackwell, Sheffield, pp 248–265
Seki M, Umezawa T, Kim JM et al. (2007) Transcriptome analysis of plant drought and salt stress response. In: Jenks MA, Hasegawa PM, Jain SM (eds) Advances in molecular breeding toward drought and salt tolerant crops. Springer Publishing, Dordrecht, pp 261–283
Shi H, Lee BH, Wu SJ et al. (2003) Overexpression of a plasma membrane Na^+/H^+ antiporter gene improves salt tolerance in *Arabidopsis thaliana*. Nat Biotechnol 21:81–85
Shinozaki K, Yamaguchi-Shinozaki K (2000) Molecular responses to dehydration and low temperature: Differences and cross-talk between two stress signaling pathways. Curr Opin Plant Biol 3:217–223
Shinozaki K, Yamaguchi-Shinozaki K, Seki M (2003) Regulatory network of gene expression in the drought and cold stress responses. Curr Opin Plant Biol 6:410–417
Simpson SD, Nakashima K, Narusaka Y et al. (2003) Two different novel *cis*-acting elements of *erd1*, a *clpA* homologous *Arabidopsis* gene function in induction by dehydration stress and dark-induced senescence. Plant J 33:259–270

Sridha S, Wu K (2006) Identification of AtHD2C as a novel regulator of abscisic responses in *Arabidopsis*. Plant J 46:124–133
Stockinger EJ, Mao Y, Regier MK et al. (2001) Transcriptional adaptor and histone acetyltransferase proteins in Arabidopsis and their interactions with CBF1, a transcriptional activator involved in cold-regulated gene expression. Nucleic Acids Res 29:1524–1533
Stolc V, Samanta MP, Tongprasit W et al. (2005) Identification of transcribed sequences in *Arabidopsis thaliana* by using high-resolution genome tiling arrays. Proc Natl Acad Sci U S A 102:4453–4458
Sunkar R, Chinnusamy V, Zhu J et al. (2007) Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. Trends Plant Sci 12:301–309
Taji T, Ohsumi C, Iuchi S et al. (2002) Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in *Arabidopsis thaliana*. Plant J 29:417–426
Taji T, Seki M, Satou M et al. (2004) Comparative genomics in salt tolerance between *Arabidopsis* and *Arabidopsis*-related halophyte salt cress using arabidopsis microarray. Plant Physiol 135:1697–1709
Teige M, Scheikl E, Eulgem T et al. (2004) The MKK2 pathway mediates cold and salt stress signaling in *Arabidopsis*. Mol Cell 15:141–152
Tokimatsu T, Sakurai N, Suzuki H et al. (2005) KaPPA-View. A web-based analysis tool for integration of transcript and metabolite data on plant metabolic pathway maps. Plant Physiol 138:1289–1300
Tran LSP, Nakashima K, Sakuma Y et al. (2004) Functional analysis of *Arabidopsis* NAC transcription factors controlling expression of erd1 gene under drought stress. Plant Cell 16:2481–2498
Tran LSP, Nakashima K, Sakuma Y et al. (2007a) Co-expression of the stress-inducible zinc finger homeodomain ZFHD1 and NAC transcription factors enhances expression of the *ERD1* gene in Arabidopsis. Plant J 49:46–63
Tran LSP, Urao T, Qin F et al. (2007b) Functional analysis of AHK1/ATHK1 and cytokinin receptor histidine kinases in response to abscisic acid, drought, and salt stress in Arabidopsis. Proc Natl Acad Sci U S A 104:20623–20628
Ulm R, Ichimura K, Mizoguchi T et al. (2002) Distinct regulation of salinity and genotoxic stress responses by Arabidopsis MAP kinase phosphatase 1. EMBO J 21:6483–6493
Umezawa T, Yoshida R, Maruyama K et al. (2004) SRK2C, a SNF1-related protein kinase 2, improves drought tolerance by controlling stress-responsive gene expression in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A 101:17306–17311
Umezawa T, Fujita M, Fujita Y et al. (2006) Engineering drought tolerance in plants: discovering and tailoring genes to unlock the future. Curr Opin Biotechnol 17:113–122
Uno Y, Furihata T, Abe H et al. (2000) Arabidopsis basic leucing zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. Proc Natl Acad Sci U S A 97:11632–11637
Urao T, Yakubov B, Satoh R et al. (1999) A transmembrane hybrid-type histidine kinase in Arabidopsis functions as an osmosensor. Plant Cell 11:1743–1754
Valliyodan B, Nguyen HT (2006) Understanding regulatory networks and engineering for enhanced drought tolerance in plants. Curr Opin Plant Biol 9:189–195
Vanderauwera S, De Block M, Van de Steene N et al. (2007) Silencing of poly(ADP-ribose) polymerase in plants alters abiotic stress signal transduction. Proc Natl Acad Sci U S A 104:15150–15155
Vinocur B, Altman A (2005) Recent advances in engineering plant tolerance to abiotic stress: achievements and limitations. Curr Opin Biotechnol 16:123–132
Vlachonasios KE, Thomashow MF, Triezenberg SJ (2003) Disruption mutations of *ADA2b* and *GCN5* transcriptional adaptor genes dramatically affect Arabidopsis growth, development, and gene expression. Plant Cell 15:626–638
Vogel JT, Zarka DG, Van Buskirk HA et al. (2005) Roles of the CBF2 and ZAT12 transcription factors in configuring the low temperature transcriptome of *Arabidopsis*. Plant J 41:195–211
Wang H, Miyazaki S, Kawai K et al. (2003) Temporal progression of gene expression responses to salt shock in maize roots. Plant Mol Biol 52:873–891

Watkinson JI, Sioson AA, Vasquez-Robinet C et al. (2003) Photosynthetic acclimation is reflected in specific patterns of gene expression in drought-stressed loblolly pine. Plant Physiol 133:1702–1716

Wong CE, Li Y, Labbe A et al. (2006) Transcriptional profiling implicates novel interactions between abiotic stress and hormonal responses in Thellungiella, a close relative of Arabidopsis. Plant Physiol 140:1437–1450

Xin Z, Browse J (1998) *eskimo1* mutants of Arabidopsis are constitutively freezing-tolerant. Proc Natl Acad Sci U S A 95:7799–7804

Xin Z, Mandaokar A, Chen J et al. (2007) Arabidopsis *ESK1* encodes a novel regulator of freezing tolerance. Plant J 49:786–799

Xiong L, Gong Z, Rock CD et al. (2001) Modulation of abscisic acid signal transduction and biosynthesis by an Sm-like protein in *Arabidopsis*. Dev Cell 1:771–781

Xiong L, Schumaker KS, Zhu JK (2002) Cell signaling during cold, drought, and salt stress. Plant Cell Suppl S165–183

Yamada K, Lim J, Dale JM et al. (2003) Empirical analysis of transcriptional activity in the *Arabidopsis* genome. Science 302:842–846.

Yamada M, Morishita H, Urano K et al. (2005) Effects of free proline accumulation in petunias under drought stress. J Exp Bot 56:1975–1981

Yamaguchi-Shinozaki K, Shinozaki K (2005) Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. Trends Plant Sci 10:88–94

Yamaguchi-Shinozaki K, Shinozaki K (2006) Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annu Rev Plant Biol 57:781–803

Yu LX, Setter TL (2003) Comparative transcriptional profiling of placenta and endosperm in developing maize kernels in response to water defecit. Plant Physiol 131:568–582

Zhang JZ (2003) Overexpression analysis of plant transcription factors. Curr Opin Plant Biol 6:430–440

Zhu JK (2002) Salt and drought stress signal transduction in plants. Annu Rev Plant Biol 53:247–273

Zhu J, Shi H, Lee BH et al. (2004) An *Arabidopsis* homeodomain transcription factor gene, *HOS9*, mediates cold tolerance through a CBF-independent pathway. Proc Natl Acad Sci U S A 101:9873–9878

Zhu J, Verslues PE, Zheng X et al. (2005) *HOS10* encodes an R2R3-type MYB transcription factor essential for cold acclimation in plants. Proc Natl Acad Sci U S A 102:9966–9971

Zimmermann P, Hirsch-Hoffmann M, Hennig L et al. (2004) GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. Plant Physiol 136:2621–2632

Chapter 15
Roles of MicroRNAs in Plant Abiotic Stress

Ricky Lewis, Venugopal Mendu, David Mcnear, and Guiliang Tang

Abstract Regulation of microRNAs (miRNAs) in response to abiotic stresses constitutes a major part of the biology of plant miRNAs. In this chapter, we exclusively review the current progress in the study of roles of miRNAs in plant abiotic stresses. The responses of various miRNAs to different types of abiotic stresses are discussed in detail. Finally, the perspective for application of the related studies and the future of plant biology are briefly presented.

15.1 Introduction

Since the discovery of *lin-4*, the first microRNA (miRNA) identified in *Caenorhabditis elegans* (Lee et al. 1993), our knowledge and understanding of the function of miRNAs in biological entities have grown tremendously. What was once thought to be a novel gene regulating mechanism important to the development in *C. elegans* has now become known as a multifunctional gene regulation system ubiquitous throughout the plant and animal kingdoms. MicroRNA, a 21 nucleotide (nt) small endogenous RNA, forms an active RNA-induced silencing complex (RISC) with a key cellular protein named Argonaute (AGO) and targets gene transcripts for degradation or translational repression. The first plant miRNAs were cloned in 2002 in *Arabidopsis thaliana* (Llave et al. 2002; Park et al. 2002; Reinhart et al. 2002). Since then, many miRNAs and their target genes have been identified and characterized in various plants species. The majority of miRNA target genes were found to encode various transcriptional factors or important functional enzymes and to play important roles in plant development and response to various biotic and abiotic stresses. In this chapter, we introduce the basic working mechanism of miRNA-directed gene regulation and discuss exclusively the role of miRNAs in abiotic stress.

R. Lewis, V. Mendu, D. Mcnear, and G. Tang (✉)
Department of Plant and Soil Sciences & KTRDC, University of Kentucky,
Cooper and University Drives, Lexington, KY, 40546-0236, USA
e-mail: guiliang.tang@uky.edu

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_15, © Springer Science+Business Media B.V. 2010

15.2 The Biogenesis of miRNAs and the Mechanism of miRNA-Directed Gene Regulation in Plants

MicroRNAs are small non-protein coding RNAs that are first transcribed by RNA polymerase II (Pol-II), usually from intergenic regions of the genome, into capped and poly (A) tailed primary transcripts termed pri-miRNAs (Tang et al. 2008). The pri-miRNAs fold up into unique stem-loop structures that can be recognized and cleaved by the Dicer-like (DCL) enzymes of the RNase III family. In the plant nucleus, DCL-1, in collaboration with its partner protein, HYL1, processes the pri-miRNAs in two steps, first into smaller stem-loop structures termed miRNA precursors (pre-miRNAs) and then into mature 21 nt long miRNA duplexes with 2 nt overhangs on the 3' end. The miRNA duplexes are transported by a nuclear transporter, HASTY, from the nucleus to the cytoplasm where they are unwound and assembled into the RISCs in single-stranded form. In plants, most miRNAs fully or extensively complement their target mRNAs, which results in a cleavage of the target mRNAs. The cleaved mRNA products are then degraded and are effectively silenced. Plant miRNAs may also regulate genes by translational repression of the target mRNAs in some cases (Chen 2004; Brodersen et al. 2008).

15.3 Plant Abiotic Stress and miRNAs

Plant abiotic stress is referred to as any adverse impact of any non-living factors on living plants in a specific environment. Unlike animals, plants are predisposed to an array of abiotic factors to which they must adapt. These abiotic factors range from required resources to pollutants. For every resource required by plants there is an optimal range for growth and reproduction. Once the availability of any of these resources reaches an upper or lower threshold we say that the plant is abiotically stressed. Thresholds are usually determined by hindrance of growth or reproduction and can vary widely amongst species. Plants may also experience abiotic stress through exposure to other environmental elements that aren't necessarily required. Soil, water, and air pollutants all have the capacity to influence plant growth if ambient levels are sufficient to inhibit or enhance plant health. Many potentially toxic elements, particularly heavy metals and salt, occur both naturally and as anthropogenic pollutants, but may be tolerated at low levels. In fact, some plants have even evolved mechanisms to tolerate, and in some cases, accumulate relatively high levels of such elements. One environmental factor that demonstrates great control over plant growth is soil pH. Soil pH greatly influences nutrient availability and high levels of H^+or OH^- can damage plant tissues. It is common for plants to be faced with more than one abiotic stress factor and when this happens, the outcome is generally an increased level of stress.

Plant abiotic stresses and response of plants to these stresses have been extensively studied. Many genes that encode both transcription factors and important detoxifying enzymes have been identified. Only recently have specific miRNAs

been identified to play important roles in various abiotic stresses. Different forms of abiotic stresses may lead to similar responses in plants during the stress. Similarly, different kinds of stresses have also been found to trigger responses in similar sets of miRNAs. This suggests that plants share common pathways involved in different abiotic stress responses. Even so, here we discuss the roles of miRNAs in each specific abiotic stress for simplicity (Table 15.1). But overlapping roles of specific miRNAs in different abiotic stresses are also discussed.

15.3.1 ABA-Mediated Responses and the miRNAs

Abscisic acid (ABA) is a multifunctional hormone that has been found in all major organs and living tissues in higher plants (Taiz and Zeiger 2002). The importance of ABA in plant stress responses is evident by its role in stomatal closure (Beardsell and Cohen 1975), maintenance of primary root growth, and inhibition of shoot growth during water stressed conditions (Saab et al. 1990). The first indication that miRNA may be involved in ABA-mediated responses came from observing ABA hypersensitivity in an *A. thaliana* mutant containing a "pleiotropic recessive Arabidopsis transposon insertion mutation," *hyl1* (Lu and Fedoroff 2000). The HYL1 protein has been identified, as the partner of DCL-1, to play a key role in miRNA biogenesis and accumulation but not in posttranscriptional regulation of transgenes, which requires a different set of RNA silencing enzymes (Han et al. 2004; Vazquez et al. 2004). These studies imply that the involvement of *HYL1* in ABA/miRNA-mediated responses lies in the production of miRNA rather than the down-stream functionality of miRNA-associated RISCs.

ABA treatment, gibberellic acid (GA) treatment, and drought stress during seed germination in *A. thaliana* have been shown to trigger the accumulation of miR159 (Reyes and Chua 2007), which was predicted to target four MYB transcription factors (MYB33, MYB65, MYB101 and MYB104) (Rhoades et al. 2002). In agreement with the prediction, Reyes and Chua (2007) found that *MYB33* and *MYB101* transcript cleavage was mediated by miR159. MYB transcription factors were found to positively regulate ABA responses by observing ABA hyposensitivity in plants with the null mutant genes *myb33* and *myb101* and in plants overexpressing miR159 (Reyes and Chua 2007). ABA induced accumulation of miR159 was found to be dependent on the ABI3 transcription factor when the *abi3-1* mutant failed to accumulate miR159 during seed germination in response to ABA treatment. Interestingly, functional redundancy was discovered in miR159a and miR159b by observing expression of the redundant gene pair *MYB33* and *MYB65* in *miR159a, miR159b*, and *miR159ab* mutants. Phenotypic abnormalities, as well as increased levels of *MYB33* and *MYB65* transcripts, were observed in *miR159ab* double mutants but not in either *miR159a* or *miR159b* single mutants. Also, transgenic plants expressing a miR159-resistant allele, *mMYB33*, were found to phenocopy the *miR159ab* double mutant, further illustrating the phenotypic consequences of miR159-mediated regulation of the MYB transcription factors (Allen et al. 2007).

Table 15.1 Abiotic stress responsive miRNAs and their functions

miRNA	Target	Target status	Function	Source
I. ABA mediated responses				
miR159	*MYB33, MYB65, and MYB101*	Confirmed	Facilitation of Anther Development	Reyes and Chua 2007
miR160	ARF10	Confirmed	Control of ABA-responsive genes	Liu et al. 2007
miR417	Unknown			
II. Oxidative stress				
miR398	*CSD1* and *CSD2*	Confirmed	Cu-Zn super oxide dismutase/free cellular Cu	Sunkar et al. 2006
III. Water deficit (drought)				
miR169	CCAAT-binding transcription factors	putative	Unknown	Zhao et al. 2007
miR393	*TIR1*	Putative	Auxin degradation and auxin-regulated transcription	Navarro et al. 2006; Zhao et al. 2007; Function of TIR1: Dharmasiri et al. 2005; Kepinski and Leyser 2005
IV. Submergence (anoxia)				
	Ill-defined			Zhang et al. 2008
V. Phosphate deficiency				
miR399	*UBC24*	Confirmed	UBC24 function is unclear	Allen et al. 2005; Bari et al. 2006
			miR399 is a long-distance Pi stress signal	Pant et al. 2008
VI. Sulfur deficiency				
miR395	ATP Sulfurylases and AST68	Confirmed	ATP sulfurylases are the first enzyme involved in the sulfur assimilation pathway	Jones-Rhoades and Bartel 2004; Function of ATP sulfurylases: Lappartient and Touraine 1996;
			AST68 is important in sulfur distribution throughout the plant	Takahashi et al. 1997

VII. Copper deficiency				
miR398	*CSD1* and *CSD2*	Confirmed	Cu-Zn super oxide dismutase/free cellular Cu	Sunkar et al. 2006
VIII. Mechanical stress responses				
miR156	SPB-like, nitrate transporter	Putative	Unknown	Lu et al. 2005
miR162	DCL-1	Putative		
miR164	NAC-domain protein	Putative		
miR408	Plastocyanin-like	5'-RACE Validated		
ptmiR475	PPR	5'-RACE Validated		
ptmiR480	Proton-dependent oligopeptide transport family protein	5'-RACE Validated		
ptmiR481	Unknown	n/a		

Studies in *A. thaliana* have also implicated a role of miR160 in ABA-mediated responses. Auxin Response Factors (ARFs), as the name implies, are transcription factors that respond to auxin, an important plant hormone involved in an array of developmental and other physiological processes. *A. thaliana* seeds and plants expressing a miR160-resistant mutant *ARF10* gene, *mARF10*, demonstrated hypersensitivity to ABA, while seeds overexpressing miR160 were less sensitive (Liu et al. 2007). Comparative transcriptome analysis between *mARF10* and *ARF10* seeds revealed that ABA-responsive genes that are generally expressed during seed maturation were overexpressed in germinating mARF10 seeds (Liu et al. 2007).

Regulation of miR417 in *A. thaliana* under ABA treatment, salt stress, and dehydration stress provides yet another example of miRNA mediated adaptive plant responses (Jung and Kang 2007). High salt stress was shown to moderately decrease transcript levels of miR417, while ABA treatment and dehydration stress generated an initial upregulation followed by downregulation (Jung and Kang 2007). When seeds overexpressing miR417 were subjected to salt treatment, germination and survival rates were decreased. Seeds overexpressing miR417 and treated with ABA demonstrated decreased germination rates when compared to WT, 40% and 80% respectively (Jung and Kang 2007). Post-germination growth was also negatively affected (Jung and Kang 2007). This miRNA is expressed in all major tissues of *A. thaliana* and is obviously influenced by multiple abiotic factors. However, the target gene(s) still remains unknown (Jung and Kang 2007).

15.3.2 Oxidative Stress and the miRNAs

Reactive Oxygen Species (ROS) are inherent to plants because they are constantly produced by aerobic processes in chloroplasts, mitochondria and peroxisomes (Apel and Hirt 2004). Elevated levels of ROS are often associated with plant stress produced by biotic factors (e.g. pathogens or herbivory) and/or abiotic factors (e.g. high light, UV radiation, temperature extremes, heavy metals, air pollutants, drought stress, salt stress, mechanical/physical stress) (Dat et al. 2000). Given that ROS may be an integral component of plant adaptive responses and potentially toxic (Dat et al. 2000), ROS concentration is under tight regulation within the plant (Dat et al. 2000; Apel and Hirt 2004).

ROS levels are maintained primarily by Superoxide Dismutases (SODs). These enzymes exist in conjunction with several co-factors including, Cu-Zn, Fe, Ni, and Mn. Cu-Zn SODs are encoded by *CSD1, CSD2*, and *CSD3* in *Arabidopsis thaliana* (Sunkar et al. 2006). The miR398 was first predicted to target *CSD1* and *CSD2* in 2004 (Bonnet et al. 2004; Jones-Rhoades and Bartel 2004). Sunkar et al. (2006) confirmed these targets and discovered that miR398 is downregulated under oxidative stress. Downregulation of miR398 is accompanied by an accumulation of *CSD1* and *CSD2* transcripts (Sunkar et al. 2006). This was shown to not be a result of stress related transcriptional induction of the Cu-Zn SOD genes, but rather, the relaxation of miR398 directed cleavage (Sunkar et al. 2006). Further confidence in the role of miR398 in oxidative stress was gained by observing the phenotypes of

A. thaliana plants expressing a miR398-resistant *CSD2* (*mCSD2*). Plants possessing *mCSD2* demonstrated greater tolerance to conditions inducing oxidative stress when compared to plants possessing the wild-type gene (Sunkar et al. 2006). Later, in Sect. 15.3.4.4, we will see that miR398 plays a crucial role in Cu homeostasis and discuss the broader implications of this finding.

15.3.3 Water Stress and the miRNAs

15.3.3.1 Water-Deficit (Drought Stress)

The somewhat spontaneous rainfall and water availability patterns experienced by most plants on earth have resulted in the evolution of adaptive responses to maintain life during times of water deficit. When water becomes scarce plants may adapt to maintain sufficient tissue hydration and/or to maintain function whilst dehydrated (Taiz and Zeiger 2002). On a grander scale, some plants have even evolved life history patterns that result in the avoidance of drought by completing their life cycles while water is abundant (Taiz and Zeiger 2002); these plants are still likely to have evolved mechanisms for dealing with water deficit within their life cycles. In this section we will be discussing the role of miRNAs in drought tolerance within the lifetime of the plant.

As discussed earlier, ABA plays a role in water stress responses. However, there are also ABA-independent response mechanisms important in drought tolerance (Yamaguchi-Shinozaki and Shinozaki 2005). In rice plants experiencing drought stress, Zhao et al. (2007) found that miR169g reached a maximum level of accumulation in roots after 6 h and in shoots after 24 h. Accompanying this finding was the identification of two ABA-independent dehydration-responsive elements up-stream of the promoter region of the gene encoding miR169g. Also temporarily upregulated was miR393, a plant miRNA thought to regulate expression of mRNAs encoding the F-box auxin receptor, *TIR1* (Navarro et al. 2006; Zhao et al. 2007). Due to its role in auxin degradation and auxin-regulated transcription, *TIR1* is important in growth and development (Dharmasiri et al. 2005; Kepinski and Leyser 2005). Thus, the findings of Zhao et al. (2007) further relate regulation of plant growth to drought tolerance.

15.3.3.2 Submergence (Anoxia)

Submergence in water can also be a source of stress in plants. The potential involvement of post-transcriptional gene regulation in submergence responses was first implied in 2003 (Dolferus et al. 2003). Submergence of plants causes low oxygen availability, which triggers adaptive strategies that alter metabolism, morphology, and physiology (Zhang et al. 2008). In maize, a total of 39 miRNAs were shown to have altered levels of expression under submergence stress (Zhang et al. 2008). Putative targets of many of these miRNAs are responsible for coding proteins that run the gamut in physiological, metabolic and morphological responses

(see Zhang et al. 2008). These putative targets were organized into three groups by the authors: those involved in organ development and plant growth, those implicated in phytohormone signal cascades, and those that encode proteins involved in diverse metabolic pathways (Zhang et al. 2008). Analysis of miRNA expression in root cells led to the construction of a potential network of known miRNAs in submergence tolerance. The findings illustrated in this study highlight the complexity of adaptive plant responses and the need to isolate regulatory factors to truly understand their function.

15.3.4 *Nutrient Scarcity and the miRNAs*

Chief among the abiotic impediments to plant growth is nutrient availability. Nutrient deficiencies may manifest themselves in a number of physiological and morphological changes. Some of these changes are due to disruptions in physiological processes while others are adaptive responses that lead to enhanced nutrient uptake, assimilation, or distribution. It is no surprise then that miRNAs are involved in nutrient acquisition in times of deprivation.

15.3.4.1 Phosphate Deficiency

The DNA and RNA that are inherent to all life on earth are constructed from nucleotides united through a phosphate bridge. Other than being a key structural component of nucleic acids and membranes, phosphate is involved in many biological functions as an active player in ATP driven energy transfer (Marschner 1995). Plants have evolved mechanisms to scavenge Pi from the soil during times of deprivation. Pi deficiency may result in changes in root architecture, such as the proliferation of roots near the topsoil, where Pi is usually found at higher concentrations (Lynch and Brown 2001), and the production of long root hairs (Bates and Lynch 2001). Pi deficiency may also trigger an increase in Pi mobilizing root exudates, mychorrhizal associations and the induction of high-affinity Pi transporters near the root surface (see Lambers et al. 2006 for an in depth review).

Most of what we know about the role of miRNA in phosphorus responses comes from studying an *A. thaliana* Pi accumulator with a mutation in *UBC24*, an ubiquitin conjugating E2 enzyme (Aung et al. 2006; Bari et al. 2006). This mutant was derived from an ethylmethane sulfonate (EMS)-mutagenized population of *A. thaliana* and was titled *pho2* because it was one of the two plants found to accumulate phosphorus (Delhaize and Randall 1995). The miR399 was computationally predicted (Jones-Rhoades and Bartel 2004), cloned, and proposed to target *UBC24* (Sunkar and Zhu 2004). This miRNA has been shown to accumulate specifically in response to Pi deprivation (Fujii et al. 2005; Bari et al. 2006). When expression of UBC24 and miR399 was observed primarily in the vascular cylinder, both were inferred to be involved in Pi shuttling (Aung et al. 2006). These findings are not entirely

consistent with the other studies that found ubiquitous expression of *UBC24* (Bari et al. 2006). The discretion likely lies in the methodology as proposed by Chiou (2007), and in spite of it, there is a clear connection between *UBC24* and miR399 expression and Pi homeostasis. *UBC24* is downregulated and miR399 is upregulated under phosphate deficiency (Fujii et al. 2005; Chiou et al. 2006); these results insinuated miR399-mediated downregulation of *UBC24*. Indeed, miR399 plays a crucial role in *UBC24* expression through RISC-mediated cleavage (Allen et al. 2005) and likely through translational repression (Bari et al. 2006). The precise mechanism by which *UBC24* influences Pi homeostasis remains unclear, but plants with the *pho2* mutation or miR399 overexpression demonstrate a significant increase in levels of a Pi transporter, *PHT1;8*, in roots (Aung et al. 2006). Chiou suggests that *UBC24* could regulate the expression of a transcription factor that is involved in Pi homeostasis (Chiou 2007).

The expression of miR399 appears to be controlled in at least two ways. It is likely that a MYB transcription factor, PHOSPHATE STARVATION RESPONSE 1 (PHR1), is involved in miR399 expression. PHR1 is expressed in response to Pi starvation and positively regulates a division of Pi-responsive genes by binding to GNATATNC *cis*-elements (Rubio et al. 2001; Franco-Zorrilla et al. 2004; Chiou 2007). This *cis*-element has been found upstream of all known miR399 genes in *A. thaliana* (Bari et al. 2006; Chiou 2007). Furthermore, *phr1* mutants demonstrate a significant decline in miR399 induction under Pi stress (Bari et al. 2006; Chiou 2007).

Expression of miR399 is also partially regulated through a process known as target mimicry. For some time the function of Pi deficiency induced non-coding RNAs found in tomato, *Arabidopsis thaliana, Medicago truncatula*, and rice (Burleigh and Harrison 1997, 1998, 1999; Liu et al. 1997; Martin et al. 2000; Wasaki et al. 2003; Shin et al. 2006) remained somewhat of a riddle. Then it was discovered that these otherwise divergent nc-RNAs, of the *IPS1/At4* family, shared a tightly conserved 24-nt sequence (Shin et al. 2006; Chiou 2007). With this discovery came the observation of mismatches between the conserved sequence and miR399 at the site of suspected cleavage (Shin et al. 2006; Franco-Zorrilla et al. 2007). The mismatch at this site not only inhibits cleavage but the mRNA sequence acts somewhat like a sponge and sequesters the miRNA/RISC (Franco-Zorrilla et al. 2007). When the mismatch is "corrected" the sequestering action of the mRNA is eradicated (Chiou 2007). Since *IPS1* is expressed in high abundance just as Pi stress is relieved, it is viewed as a mechanism for the plant to return back to homeostatic levels of miR399, its target genes, and thus, Pi.

15.3.4.2 MicroRNA as a Long Distance Signal in Phosphate Homeostasis

Long distance signals are the result of the obvious need for communication between cells that are located in separate parts of the plant body. Plants have evolved complex systems to regulate the trafficking of macromolecular signals through the plasmodesmata (Lough and Lucas 2006). Before the confirmation of miR399 as a long distance signal involved in Pi homeostasis (Pant et al. 2008), the potential

involvement of miRNA and other nc-RNAs in such networks was being seriously considered due to their existence in phloem (Lough and Lucas 2006).

Grafting experiments where either the rootstocks or scions of a mutant overexpressing miR399 were coupled with WT complementary parts were used to study the potential long-distance trafficking of the miRNA (Pant et al. 2008). Chimeric plants overexpressing miR399 in the scions demonstrated higher levels of miR399 in roots when compared to WT plants (Pant et al. 2008). When WT scions were coupled with overexpressing roots, there was no significant difference in miR399 expression in the scion (Pant et al. 2008). These results coupled with the miR399 accumulation in the roots of Pi starved plants indicated a predominantly unidirectional flow of miR399 from the shoot to the root. Observation of *PHO2 (UBC24)* cleavage products indicated that miR399 is likely functioning in the roots (Pant et al. 2008). In short, Pi starvation induces the production of miR399 in shoots; miR399 is then shuttled to the roots where it alters the expression of *UBC24* as described above.

15.3.4.3 Sulfur Deficiency

Sulfur is an essential macronutrient that is a key component of cysteine and methionine amino acids that may become incorporated into S-containing coenzymes and secondary plant products (Marschner 1995). In fact, cysteine is a precursor to glutathione, a tripeptide that is also a powerful antioxidant which plays a key function in protecting chloroplasts from oxidative damage (Marschner 1995). MiR395 has been found to be involved in S metabolism by targeting both ATP sulfurylases and the sulfate transporter AST68 (Jones-Rhoades and Bartel 2004; Jones-Rhoades et al. 2006). ATP sulfurylases are the first enzyme involved in the sulfur assimilation pathway and increase in expression in response to sulfur deprivation (Lappartient and Touraine 1996). The sulfate transporter AST68 is important in sulfur distribution throughout the plant and is induced under S starvation (Takahashi et al. 1997). Unlike the downregulation of the target gene SODs mediated by the upregulation of miR398 miR395 was induced along with the target genes (Jones-Rhoades and Bartel 2004), which represents somewhat of a conundrum. Nevertheless, it may suggest an auto/co-regulation and/or homestasis between miR395 and its target genes, which are involved in both transcriptional and post-transcriptional regulations. Similar observations for other miRNAs and their targets were also found in tomato fruit development (Mendu et al. 2009, manuscript submitted). Further investigation is required to understand the nuances of how the expression of miRNAs and their target genes were not adversely correlated and the mechanisms by which the transcription and post-transcription of miRNAs and their target genes are controlled.

15.3.4.4 Copper Deficiency

Copper is an essential plant micronutrient important in photosynthesis, oxidative responses and more (Marschner 1995). Cu deficiency induces miR398, which then

negatively regulates the translation of *CSD1* and *CSD2* into Cu-SOD proteins (Yamasaki et al. 2007). In higher plants the Cu/Zn-SODs are replaced by Fe-SODs (Quinn and Merchant 1995). Reduction of Cu containing SOD proteins increases the availability of Cu to be used in other biological processes, one of which is plastocyanin function (Marschner 1995). On the other hand, high concentration of Cu is toxic to plants. Arabidopsis plants overexpressing miR398 were more sensitive to Cu (V.M. and G. T. unpublished data). Because miR398 partially mediates the level of free Cu in the plant, it could be said that it is a part of the regulatory network that controls the Cu ionome.

15.3.5 Mechanical Stress Responses and the miRNAs

Plants experience mechanical stress when bombarded by wind, water, or any other entity that imposes physical force upon the plant body. Studies concerning miRNA and mechanical stress are still in their infancy; however, some interesting results have come from observing miRNA expression in tension- and compression-stressed *Populus trichocarpa*. The following results were obtained by performing comparative analysis of miRNA expression in *P. trichocarpa* experiencing mechanical stress via bending the plant stem in an arch for 4 days and in plants free of mechanical stress. *P. trichocarpa* miRNA transcript levels of miR156, miR162, miR164, miR475, miR480, and miR481 were suppressed in the xylem tissue of mechanically stressed plants when compared to the un-stressed *P. trichocarpa* (Lu et al. 2005). Relative induction of transcript levels of ptr-miR408 were observed in stressed xylem tissues (Lu et al. 2005). The study of mutants overexpressing specific miRNAs or miRNA-resistant target sequences would aid in determining the role of these miRNAs in mechanical stress responses. The authors of this study noted the importance of identifying conserved and tree specific miRNAs that were mediated by mechanical stress. Many conserved miRNAs have evolved different functions, and the existence of tree specific miRNAs reveals the evolution of regulatory devices that are associated with divergence. It is thought that studying the evolution of these miRNAs and their functions will enrich our understanding of gene regulatory changes associated with trait emergence and speciation (Lu et al. 2005).

15.3.6 Dynamic Regulation of miRNAs in Response to Abiotic Stresses

Various studies have identified numerous miRNAs that were either up regulated or down regulated upon stress treatment (see Table 15.1 for a summary). Recent studies indicate that the responses of miRNAs to various abiotic stresses are more dynamic than previously appreciated. In maize, the expressions of miR159, miR395, miR474 and miR528 were suppressed at the early stage of water submergence and increased

after 24 h post-submergence, showing a dynamic regulation of miRNAs during water stress (Zhang et al. 2008).

More recently, we showed that the regulation of miR398 in response to ABA and salt stress was more dynamic in plants than previously reported (Jia et al. 2009). Our results demonstrated that, in poplars, miR398 was first induced upon 3–4 h of ABA or salt stress. However, this induction was declined after 48 h and finally accumulated again over a longer stress (72 h). In contrast, such dynamic regulation of miR398 under salt stress was completely lost in Arabidopsis in which miR398 was steadily suppressed. Interestingly, ABA treatment caused a deviate dynamic regulation of miR398 in Arabidopsis, showing an opposite response as compared to that in poplars. Furthermore, the expression of miR398 target genes, superoxide dismutase (SOD) or Cu/Zn superoxide dismutases (CSDs), was in perfect reverse correlation with the miR398 level, suggesting a primary control of SOD or CSD expression by miR398 under abiotic stress. Taken together, our data consistently show a correlated dynamic regulation between miR398 and its targets by ABA and salt stress, and raise the possibility that miRNAs might be dynamically and differentially regulated under various stress conditions in different plant species.

15.4 Other Small RNAs and Abiotic Stress

Other endogenous small RNAs, such as short interfering RNAs (siRNAs), are also involved in RISC mediated gene regulation much like miRNAs, but differ in their biogenesis (Tang 2005). Although there are a variety of ways that siRNAs can be generated in plants, they all share the common feature of a relatively long dsRNA precursor strand from which a relatively short sequence of RNA (21–26 nt) is produced. Some siRNAs are derived from dsRNAs produced from transcripts of repeat DNA and heterochromatin (Mallory and Vaucheret 2006); others are generated from dsRNA formed from the products of miRNA-directed cleavage of mRNA (Vaucheret 2005; Vazquez et al. 2004). The siRNAs may also be derived from dsRNAs produced by the annealing of two antisense mRNAs encoded by natural *cis*-antisense gene pairs; these siRNAs are termed nat-siRNAs (Borsani et al. 2005).

SiRNA-mediated stress responses were first observed in salt stressed *A. thaliana* (Borsani et al. 2005). A siRNA generated from the 3' end of SRO5 (At5g62520) was found to be complementary to the 3' end of Δ1-pyroline-5-carboxylate dehydrogenase, known as P5CDH (Borsani et al. 2005). SRO5 is salt stress induced and the transcripts complement with *P5CDH* transcripts to form a dsRNA that is the precursor for a 24-nt nat-siRNA that mediates the cleavage of *P5CDH* transcripts (Borsani et al. 2005). A 21-nt nat-siRNA is derived from the cleavage products, and this nat-siRNA may mediate the cleavage of more *P5CDH* transcripts (Borsani et al. 2005). Downregulation of P5CDH is accompanied by accumulation of proline, which is thought to aid in stress tolerance by ROS scavenging and possibly osmoprotection (Borsani et al. 2005).

15.5 MicroRNA, Abiotic Stress, and the Future of Plant Biology

Plants are constantly subjected to various abiotic stresses and miRNAs are likely acquiring new functions in regulating gene expression in response to these stresses. By focusing effort on the study of plant abiotic stress responses, we are gaining insights into plant biology that can be applied practically in crop systems and to unlock the secrets of life. As mentioned earlier, the expanding human population is increasingly stressing agroecosystems. Farming practices throughout the world have led to the degradation of soil quality, and many soils have a naturally low occurrence of plant required and/or beneficial soil components. Furthermore, these abiotic stresses are changing in both type and degree with intensive industrial and farming practices. Previously dominant types of abiotic stresses may disappear and new types of abiotic stresses may emerge. For example, zinc is becoming more apparent as a "trouble nutrient" because much of the world exhibits moderate to intense Zn deficiency in soils due to farming and natural occurrence, and this is often accompanied by Zn deficiency in humans (Cakmak 2008). Nevertheless, it is almost completely unknown how zinc deficiency affects the expression of miRNAs or whether it induces birth of new miRNAs. Identification of new abiotic stress regulated miRNAs would help understand the impact of abiotic stresses on plant productivity.

On the other hand, certain abiotic stresses are becoming more severe. For example, UV-B stress becomes evident with the gradual destruction of the ozone layer. A number of miRNAs were induced or suppressed by UV-B treatment in Populus tremula (Jia et al., 2009). These could serve as effective environmental biomarkers for monitoring the change of UV-B intensity or the integrity of the ozone layer. Most importantly, in depth understanding of the roles of these abiotic stress-regulated miRNAs and their target genes will eventually contribute to the improvement of plant productivity to feed an increased population of human beings.

Application of miRNA analysis and experimentation could also help reveal ecophysiological responses to abiotic stress in natural ecosystems. Climate change and pollution have already begun to alter the ecosystem functioning. Understanding the genes responsible for plant adaptation to the environment will help us predict future changes and protect species at particular risk. It is quite easy to imagine ecosystem management/ecological research that involves plants expressing different levels of endogenous miRNA or artificial miRNA under various environmental scenarios and extrapolates the level of tolerance or vulnerability conferred by the phenotypic differences. In conclusion, miRNAs are not only key players in plant abiotic stress responses, but also could be employed as a tool (artificial miRNAs) to manipulate the expression of many identified and to be identified stress-response genes. In the coming years, miRNA biotechnology will be instrumental in expanding our understanding of the genetic aspects of every avenue of plant biology.

Acknowledgements G.T. is supported by the Kentucky Tobacco Research and Development Center (KTRDC), the USDA-NRI grants 2006-35301-17115 and 2006-35100-17433, the NSF

MCB-0718029 (Subaward No. S-00000260), and an award from the Kentucky Science and Technology Corporation, under the Contract # KSTC-144-401-08-029.

References

Allen E, Xie ZX, Gustafson AM, Carrington JC 2005 MicroRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell 121:207–221

Allen RS, Li JY, Stahle MI, Dubroue A, Gluber F, Millar AA 2007 Genetic analysis reveals functional redundancy and the major target genes of the Arabidopsis miR159 family. Proc Natl Acad Sci USA 104:16371–16376

Apel K, Hirt H 2004 Reactive oxygen species: metabolism, oxidative stress, and signal transduction. Annu Rev Plant Biol 55:373–399

Aung K, Lin SI, Wu CC, Huang YT, Su CL, Chiou TJ 2006 pho2, a phosphate overaccumulator, is caused by a nonsense mutation in a microRNA399 target gene. Plant Physiol 141:1000–1011

Bari R, Pant BD, Stitt M, Scheible WR 2006 PHO2, microRNA399, and PHR1 define a phosphate-signaling pathway in plants. Plant Physiol 141:988–999

Bates TR, Lynch JP 2001 Root hairs confer a competitive advantage under low phosphorus availability. Plant Soil 236:243–250

Beardsell MF, Cohen D 1975 Relationships between leaf water status, abscisic-acid levels, and stomatal resistance in maize and sorghum. Plant Physiol 56:207–212

Bonnet E, Wuyts J, Rouze P, Van de Peer Y 2004 Detection of 91 potential in plant conserved plant microRNAs in Arabidopsis thaliana and Oryza sativa identifies important target genes. Proc Natl Acad Sci USA 101:11511–11516

Borsani O, Zhu JH, Verslues PE, Sunkar R, Zhu JK 2005 Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. Cell 123:1279–1291

Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, Sieburth L, Voinnet O 2008 Widespread Translational inhibition by plant miRNAs and siRNAs. Science 320:1185–1190

Burleigh SH, Harrison MJ 1997 A novel gene whose expression in Medicago truncatula roots is suppressed in response to colonization by vesicular-arbuscular mycorrhizal (VAM) fungi and to phosphate nutrition. Plant Mol Biol 34:199–208

Burleigh SH, Harrison MJ 1999 The down-regulation of Mt4-like genes by phosphate fertilization occurs systemically and involves phosphate translocation to the shoots. Plant Physiol 119:241–248

Burleigh SM, Harrison MJ 1998 Characterization of the Mt4 gene from Medicago truncatula. Gene 216:47–53

Cakmak I 2008 Enrichment of cereal grains with zinc: agronomic or genetic biofortification? Plant Soil 302:1–17

Chen X 2004 A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. Science 303:2022–2025

Chiou TJ 2007 The role of microRNAs in sensing nutrient stress. Plant Cell Environ 30:323–332

Chiou TJ, Aung K, Lin SI, Wu CC, Chiang SF, Su CL 2006 Regulation of phosphate homeostasis by microRNA in Arabidopsis. Plant Cell 18:412–421

Dat J, Vandenabeele S, Vranova E, Van Montagu M, Inze D, Van Breusegem F 2000 Dual action of the active oxygen species during plant stress responses. Cell Mol Life Sci 57:779–795

Delhaize E, Randall PJ 1995 Characterization of a phosphate-accumulator mutant of Arabidopsis-Thaliana. Plant Physiol 107:207–213

Dharmasiri N, Dharmasiri S, Estelle M 2005 The F-box protein TIR1 is an auxin receptor. Nature 435:441–445

Dolferus R, Klok EJ, Delessert C, Wilson S, Ismond KP, Good AG, Peacock WJ, Dennis ES 2003 Enhancing the anaerobic response. Ann Bot Lond 91:111–117

Franco-Zorrilla JM, Gonzalez E, Bustos R, Linhares F, Leyva A, Paz-Ares J 2004 The transcriptional control of plant responses to phosphate limitation. J Exp Botany 55:285–293
Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, Garcia JA, Paz-Ares J 2007 Target mimicry provides a new mechanism for regulation of microRNA activity. Nat Genet 39:1033–1037
Fujii H, Chiou TJ, Lin SI, Aung K, Zhu JK 2005 A miRNA involved in phosphate-starvation response in Arabidopsis. Curr Biol 15:2038–2043
Han MH, Goud S, Song L, Fedoroff N 2004 The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. Proc Natl Acad Sci USA 101:1093–1098
Jia X, Ren L, Chen QJ, Li R, Tang G 2009 UV-B responsive microRNAs in Populus tremula. J of Plant Physiol (in press)”
Jia X, Wang, WX, Ren L, Chen QJ, Mendu V, Willcut B, Dinkins R, Tang X, Tang G 2009 Differential and dynamic regulation of miR398 and its targets in response to ABA and salt stress in Populus tremula and Arabidopsis thaliana. Plant Mol Biol 71: 51–59
Jones-Rhoades MW, Bartel DP 2004 Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol Cell 14:787–799
Jones-Rhoades MW, Bartel DP, Bartel B 2006 MicroRNAs and their regulatory roles in plants. Annu Rev Plant Biol 57:19–53
Jung HJ, Kang H 2007 Expression and functional analyses of microRNA417 in Arabidopsis thaliana under stress conditions. Plant Physiol Biochem 45:805–811
Kepinski S, Leyser O 2005 The Arabidopsis F-box protein TIR1 is an auxin receptor. Nature 435:446–451
Lambers H, Shane MW, Cramer MD, Pearse SJ, Veneklaas EJ 2006 Root structure and functioning for efficient acquisition of phosphorus: matching morphological and physiological traits. Ann Bot Lond 98:693–713
Lappartient AG, Touraine B 1996 Demand-driven control of root ATP sulfurylase activity and SO42-uptake in intact canola - The role of phloem-translocated glutathione. Plant Physiol 111:147–157
Lee RC, Feinbaum RL, Ambros V 1993 The C-Elegans Heterochronic Gene Lin-4 encodes small RNAs with Antisense Complementarity to Lin-14. Cell 75:843–854
Liu CM, Muchhal US, Raghothama KG 1997 Differential expression of TPS11, a phosphate starvation-induced gene in tomato. Plant Mol Biol 33:867–874
Liu PP, Montgomery TA, Fahlgren N, Kasschau KD, Nonogaki H, Carrington JC 2007 Repression of AUXIN RESPONSE FACTOR10 by microRNA160 is critical for seed germination and post-germination stages. Plant J 52:133–146
Llave C, Kasschau KD, Rector MA, Carrington JC 2002 Endogenous and silencing-associated small RNAs in plants. Plant Cell 14:1605–1619
Lough TJ, Lucas WJ 2006 Integrative plant biology: role of phloem long-distance macromolecular trafficking. Annu Rev Plant Biol 57:203–232
Lu C, Fedoroff N 2000 A mutation in the Arabidopsis HYL1 gene encoding a dsRNA binding protein affects responses to abscisic acid, auxin, and cytokinin. Plant Cell 12:2351–2365
Lu SF, Sun YH, Shi R, Clark C, Li LG, Chiang VL 2005 Novel and mechanical stress-responsive microRNAs in Populus trichocarpa that are absent from Arabidopsis. Plant Cell 17:2186–2203
Lynch JP, Brown KM 2001 Topsoil foraging - An architectural adaptation of plants to low phosphorus availability. Plant Soil 237:225–237
Mallory AC, Vaucheret H 2006 Functions of microRNAs and related small RNAs in plants. Nat Genet 38:S31–S36
Marschner H 1995 Mineral nutrition of higher plants, 2nd edn. Academic, London
Martin AC, del Pozo JC, Iglesias J, Rubio V, Solano R, de la Pena A, Leyva A, Paz-Ares J 2000 Influence of cytokinins on the expression of phosphate starvation responsive genes in Arabidopsis. Plant J 24:559–567
Navarro L, Dunoyer P, Jay F, Arnold B, Dharmasiri N, Estelle M, Voinnet O, Jones JDG 2006 A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. Science 312:436–439

Pant BD, Buhtz A, Kehr J, Scheible WR 2008 MicroRNA399 is a long-distance signal for the regulation of plant phosphate homeostasis. Plant J 53:731–738

Park W, Li JJ, Song RT, Messing J, Chen XM 2002 CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. Curr Biol 12:1484–1495

Quinn JM, Merchant S 1995 2 Copper-responsive elements associated with the Chlamydomonas Cyc6 gene-function as targets for transcriptional activators. Plant Cell 7:623–638

Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP 2002 MicroRNAs in plants. Genes Dev 16:1616–1626

Reyes JL, Chua NH 2007 ABA induction of miR159 controls transcript levels of two MYB factors during Arabidopsis seed germination. Plant J 49:592–606

Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, Bartel DP 2002. Prediction of plant microRNA targets. Cell 110:513–520

Rubio V, Linhares F, Solano R, Martin AC, Iglesias J, Leyva A, Paz-Ares J 2001 A conserved MYB transcription factor involved in phosphate starvation signaling both in vascular plants and in unicellular algae. Genes Dev 15:2122–2133

Saab IN, Sharp RE, Pritchard J, Voetberg GS 1990 Increased endogenous abscisic-acid maintains primary root-growth and inhibits shoot growth of maize seedlings at low water potentials. Plant Physiol 93:1329–1336

Shin H, Shin HS, Chen R, Harrison MJ 2006 Loss of At4 function impacts phosphate distribution between the roots and the shoots during phosphate starvation. Plant J 45:712–726

Sunkar R, Zhu JK 2004 Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. Plant Cell 16:2001–2019

Sunkar R, Kapoor A, Zhu JK 2006 Posttranscriptional induction of two Cu/Zn superoxide dismutase genes in Arabidopsis is mediated by downregulation of miR398 and important for oxidative stress tolerance. Plant Cell 18:2051–2065

Taiz La, Zeiger E 2002 Plant physiology, 3rd edn. Sinauer Associates, Sunderland, MA

Takahashi H, Yamazaki M, Sasakura N, Watanabe A, Leustek T, Engler JD, Engler G, VanMontagu M, Saito K 1997 Regulation of sulfur assimilation in higher plants: a sulfate transporter induced in sulfate-starved roots plays a central role in Arabidopsis thaliana. Proc Natl Acad Sci USA 94:11102–11107

Tang G, Tang X, Mendu V, Jia X, Chen QJ, He L 2008 The art of microRNA: various strategies leading to gene silencing via an ancient pathway. Biochim Biophys Acta 1779:655–662

Tang GL 2005 siRNA and miRNA: An insight into RISCs. Trends Biochem Sci 30:106–114

Vaucheret H 2005 MicroRNA-dependent trans-acting siRNA production. Sci STKE 2005(300): pe43

Vazquez F, Gasciolli V, Crete P, Vaucheret H 2004 The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. Curr Biol 14:346–351

Wasaki J, Yonetani R, Shinano T, Kai M, Osaki M 2003 Expression of the OsPI1 gene, cloned from rice roots using cDNA microarray, rapidly responds to phosphorus status. New Phytol 158:239–248

Yamaguchi-Shinozaki K, Shinozaki K 2005 Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. Trends Plant Sci 10:88–94

Yamasaki H, Abdel-Ghany SE, Cohu CM, Kobayashi Y, Shikanai T, Pilon M 2007 Regulation of copper homeostasis by micro-RNA in Arabidopsis. J Biol Chem 282:16369–16378

Zhang Z, Wei L, Zou X, Tao Y, Liu Z, Zheng Y 2008 Submergence-responsive microRNAs are potentially involved in the regulation of morphological and metabolic adaptations in maize root cells. Ann Bot 102:509–519

Zhao BT, Liang RQ, Ge LF, Li W, Xiao HS, Lin HX, Ruan KC, Jin YX 2007 Identification of drought-induced microRNAs in rice. Biochem Biophys Res Commun 354:585–590

Chapter 16
Molecular Tools for Enhancing Salinity Tolerance in Plants

Jesus Cuartero, Maria C. Bolarin, Vicente Moreno, and Benito Pineda

Abstract Salinity is nowadays considered one of the main factors that limit crop productivity and a threat to world's food production. Hence, to breed salt tolerant varieties of crops and horticultural species is necessary to increase or at least maintain food production in order to feed the growing world's population. Plant tolerance to salinity is a complex phenomenon at both cellular and plant level. Since salt causes several types of stresses, plants face salinity using different strategies, whose relative importance depends on the species and the growing conditions. Here we present an overview of the salt tolerance mechanisms to counteract osmotic, ionic and oxidative stress, as well as an update of the knowledge on the processes involved in salt tolerance gained in part through the new genomic approaches. To breed new cultivars able to grow and maintain crop productivity on saline conditions requires variability for some of the traits related to salinity tolerance, the discovery of quantitative trait loci (QTL) regulating those traits, a deep understanding of QTL interaction with other QTL and with the environment, and the transfer of QTL from donors to elite lines using phenotypic and marker assisted selection. We have summarised part of the information related to these four issues and some guidance is given to maximize the efficiency of the selection processes. Genetic transformation has become a powerful tool in plant breeding programs since it allows the introduction of gene(s) controlling traits without affecting the rest of the characteristics of an elite genotype. In this chapter we have reviewed the available information on several topics such as: salt tolerance improvement aided by genetic transformation, functional analysis of genes related salt-tolerance, the complexity of the trait and its evaluation method, the number of genes to be introduced, and the sources of genetic variability. Finally, the use of genomic tools like transcriptomic analysis, post-transcriptional gene silencing, insertional mutagenesis and gene traps, to perform the genetic dissection of this complex trait is discussed.

J. Cuartero (✉)
Estacion Experimental La Mayora, CSIC, E-29750 Algarrobo, Málaga, Spain
e-mail: cuartero@eelm.csic.es

M.C. Bolarin
CEBAS, CSIC, Campus de Espinardo, E-30100 Espinardo, Murcia, Spain

V. Moreno and B. Pineda
IBMCP, CSIC-UPV, Avda. de los Naranjos s/n, E-46022 Valencia, Spain

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_16, © Springer Science+Business Media B.V. 2010

16.1 Introduction

Over 6% of the world's total land area (840 million hectares) is affected by salinity or sodicity (FAO 2005). This enormous area has only had small consequences for agriculture because there was enough good land for rain-fed crops: of the 1,500 million hectares dedicated to rain-fed agriculture only 32 millions were affected by secondary salinisation, the salinisation caused by agricultural practices (FAO 2005). This scenario is completely different when irrigated land is considered. The 230 million hectares of irrigated land produce about one third of world's food and over 20% of this productive land is affected by secondary salinisation (FAO 2005) because salts dissolved in the irrigation water are deposited in the soil following evapotranspiration. And the greatest danger comes from the continuously increasing secondary salinisation in irrigated areas. In California for example, where irrigation started about 130 years ago, half of the irrigated area was affected by salinity more than 20 years ago (Lewis 1984). The risk of secondary salinisation is proportional to the amount of salts in the irrigation water. This risk will increase as population increases because cities (including gardens and sport premises) and industry will pay for the best quality water leaving the worst to agriculture.

Salinity is nowadays considered one of the main factors that limit crop productivity and a threat to world's food production. In addition, world's population is expected to increase about 50% by 2050. Hence, it is necessary to sustain agricultural productivity on the increasingly saline irrigated land by salt tolerant cultivars. However, it should be pointed out that salt-tolerant cultivars will only be part of the solution. Salinity in agricultural systems tends to increase rather than remain stable, and may rise in the 8–10 years period that takes to release a new cultivar. For this reason, plant breeding cannot be the full answer to the problem of salinisation but part of an integrated programme that would also include irrigation and drainage management (Flowers and Yeo 1995).

Higher plants do not have a salt-tolerant metabolism even if the plant itself thrives in seawater (Flowers 1972a, b). Salt tolerance in organisms other than the halobacteria do not depend on salt-tolerant proteins, but on keeping a defined micro-environment in the cytoplasm, regulated for the quantity and quality of inorganic ions. Single-celled aquatic organisms or cell cultures can excrete into the medium any dangerous substance. Animals have their cells designed to excrete into the bloodstream and special organs to deal with these toxic compounds. The plant root can only excrete into the solution being drawn towards it by transpiration, but the aboveground cells have nowhere to excrete apart from the small volume of the apoplast. The large central vacuole is the limited place where plant cells can store toxic materials, hence its importance in saline conditions. Glicophytes, where most of the crops are included, are not affected by the external salinity per se, but by growth fall below net ion import leading to ever-increasing internal concentrations and, eventually, to catastrophic failure (Munns and Termaat 1986). Contrarily, in halophytes, growth and net salt uptake are coupled, even if we do not know what controls what (Yeo and Flowers 1986). The targets to produce a

plant tolerant to salinity are ion transport and compartmentalization and the synthesis of compatible solutes (osmolytes or osmoprotectants) to counteract osmotic and ionic stresses, oxidative protection, and the maintenance of ion homeostasis. But plant responses to salinity are even more complex at both the whole plant level and cellular level, and research efforts have been focused on understanding the physiological and molecular basis of salt tolerance in higher plants (Mansour and Salama 2004; Sairam and Tyagi 2004; Bartels and Dunkar 2005; Munns and Tester 2008).

In the first section of this chapter, we will describe the mechanisms developed by plants to cope with salinity and the relative importance of each one of them, which vary not only with the species but also with other factors such as stress level, exposure time to salinity, environmental conditions, etc. Advances in our knowledge of the physiological and molecular basis of salt tolerance so far achieved, with special mention to the emerging progress thanks to the new genomic approaches, are summarised. The achievement of salt tolerant cultivars is reviewed in the second part of the chapter, with special attention to the search for natural variation in phenotypic and physiological characters, the discovery of quantitative trait loci regulating salt tolerance, the interaction of these loci with the environment and other loci, and to design efficient breeding programmes based in marker assisted selection coupled with phenotypic selection. In the next section, the transfer and expression of genes related to salt tolerance into elite cultivars or parents of modern F_1 hybrids through genetic transformation is a very attractive approach because, this way, susceptible but productive cultivars could be transformed into tolerant ones, while maintaining all the valuable traits today's cultivars possess. Here we will review the progress in this field, raise the problems associated with the functional analysis of salt tolerance-related genes and discuss the adequacy of genomic tools to perform the genetic dissection of this complex trait.

16.2 Salt Tolerance Mechanisms at Physiological and Molecular Levels

Salt causes several types of plant stress including osmotic stress and ionic stress due to the accumulation of toxic saline ions, nutritional stress due to the altered nutrient uptake, especially of ions such as K^+ and Ca^{2+}, and oxidative stress. However, the main physiological mechanism responsible of plant growth reduction and how environment and genotype modulate plant salinity response are not yet known.

To understand the physiological mechanisms responsible for salinity tolerance, it is important to bear in mind that plants respond to salinity by using two main tolerance mechanisms: mechanisms of osmotic tolerance to avoid the osmotic effect of the salt outside the roots, which occurs normally at low stress levels and over short periods of salt stress, and mechanisms of ionic tolerance to avoid the

toxic effect of the salt within the plant, the main effect induced by long-term or high stress levels. However, the timescale over which ion-specific damage is manifested depends on the salt sensitivity of the species and the stress level. Thus, the ion-specific effect will start earlier in plants with a low ability to regulate the transport of saline ions to the shoot, or when high salt levels are applied. Although the two phases are generally separated in time for most plants, it is also possible for ion toxicity to take effect during the first phase itself and for osmotic effects to persist in the second phase (De Costa et al. 2007; Muñoz-Mayor et al. 2008).

16.2.1 Plant Response to Osmotic Stress Induced by Salinity

The osmotic effect starts immediately after the imposition of salt stress, before the saline ions are taken up by the roots. The salt concentration around the roots reduces the osmotic potential of the growing medium and thus the plant's ability to take up water and nutrients by the roots, similar to the effect induced by drought. (Munns 2002). The plant needs to accumulate solutes to maintain cell volume and turgor, so the response to turgor reduction is osmotic adjustment, which is a major component of salt stress tolerance. Osmotic potential may be reduced either by the simple effect of solute concentration due to reduced water uptake by the plant or by the active solute accumulation. Osmotic adjustment is defined as the lowering of osmotic potential in plant tissue due to net accumulation of solute (Blum et al. 1996).

The main solutes contributing to osmotic adjustment are the inorganic solutes, which are taken up from the substrate and transported to the shoot, and the organic solutes, which are synthesised by the plant. Na^+ and Cl^- are energetically efficient osmolytes for osmotic adjustment, but must be compartmentalised into the vacuole to minimise cytotoxicity. Within the cytoplasm, osmotic adjustment is achieved by accumulation of compatible osmolytes. Some compatible osmolytes are essential elemental ions, such as K^+, but the majority of these are organic solutes, especially sugars (mainly fructose and glucose) and organic acids. Other osmolytes are sugar alcohols (glycerol and methylated inositols), complex sugars (trehalose, raffinose and fructans), quaternary amino acid derivatives (proline, glycine betaine, β-alanine betaine, proline betaine), tertiary amines and sulfonium compounds (Zhu 2001). Another function of compatible osmolytes that may occur at lower concentrations is osmoprotection, which includes protection of thylakoid and plasma membrane integrity, stabilising proteins, a sink for energy or reducing power, a source of carbon and nitrogen for recovery, or scavenging of reactive oxygen species generated by salt stress (Bartels and Dunkar 2005). These beneficial impacts have been shown by overexpression of different genes involved in osmolyte accumulation (Ge et al. 2008; Yang et al. 2008). However, consensus has not been reached on the effectiveness of accumulation of osmolytes by genetic modification, as occurs in the case of proline. Therefore, further studies are required to give an insight into the role of the osmolytes and their effective utilisation for stress tolerance.

The use of organic solutes for osmotic adjustment is energetically much more expensive than the use of the saline ions proceeding from the substrate (Yeo 1983). Thus, the ATP requirement for the synthesis or accumulation of solutes in leaves was reported by Ravens (1985) as 3.5 for Na^+, 34 for mannitol, 41 for proline, 50 for glycine betaine and about 52 for sucrose. In this respect, salt tolerance may not always be associated with low Na^+ concentration in the leaves. While the more tolerant genotypes of many species are those better able to prevent excessive ion accumulation, the leaves of halophytes do contain high salt concentrations (Santa-Cruz et al. 1999), which are necessary to adjust the leaf water relations to low external potentials and plants use the cheapest solutes from an energetic point of view (Neumann 1997). Such a situation also seems to occur in species that are not very sensitive to salt, such as tomato grown at low-mid salinity levels, as the higher the fruit yield, the higher the contribution of inorganic solutes, including the saline ions, to the osmotic potential (Alarcon et al. 1993; Estañ et al. 2005). Thus, breeding for Na^+ accumulation, rather than exclusion, could be a more effective strategy for improving salt tolerance of conventional crop plants.

16.2.2 Plant Response to Ionic Stress Induced by Salinity

Since NaCl is the major component of most saline soils, our usage of the terms salinity and salt stress here refers to stress caused by NaCl. The ionic stress starts generally later than osmotic stress, when the toxic saline ions are transported to the shoot and build up to toxic levels within the leaves. Thus, roots must exclude most of the Na^+ and Cl^- dissolved in the growing medium, or the salt in the shoot will gradually build up with time to toxic levels. The cause of injury is probably salt load exceeding the ability of the cells to compartmentalise salts in the vacuole. Salts would then build up rapidly in the cytoplasm and inhibit enzyme activity (Yokoi et al. 2002). Alternatively, they might build up in the cell walls and dehydrate the cell (Flowers et al. 1991). The rate of cell death is crucial for the survival of the plant. If new leaves are continually produced at a rate greater than that at which old leaves die, there will be enough photosynthesizing leaves on the plant to produce flowers and seeds, although reduced in number. However, if old leaves die more quickly than new ones develop, the plant may not survive.

In most studies on salinity it has not been possible to determine whether the toxic effects observed are due to Na^+, Cl^- or to a contribution of the two. In only a few species such as citrus (Moya et al. 2003) has there been conclusive evidence of greater sensitivity to Cl^- than to Na^+. Consequently, Na^+ is considered the primary cause of ion-specific damage for many plants (Tester and Davenport 2003). However, similar relationships between fruit yield and leaf ionic concentrations for Na^+ and Cl^- were observed in tomato, which suggests that the toxic effects are due, at least in the long term, to the contribution of both ions (Estañ et al. 2005).

Salt tolerance of cultivated species is generally correlated to an efficient Na^+ and Cl^- exclusion mechanism and better maintenance of leaf K^+ concentration at

high external NaCl (Gorham et al. 1990; Matsushita and Matoh 1991). However, the higher salt tolerance of wild tomato species over cultivated forms has generally been associated with the halophytic character of Na^+ accumulation in the wild relatives (Cuartero and Fernandez-Muñoz 1999). Tester and Davenport (2003) suggested that although halophytes accumulate Na^+ in the shoot, it is unlikely that halophytic species have higher rates of Na^+ transport at high salinities or over the long term than salt-sensitive species. Studies of halophytes at low salinities tend to obscure the true situation, because many halophytes show growth stimulation upon addition of NaCl to a growth medium when NaCl is rapidly accumulated and employed preferentially as an osmoticum (Alarcon et al. 1993; Glenn et al. 1999). Several recent reviews of this area of study have been carried out (Tester and Davenport 2003; Apse and Blumwald 2007; Munns and Tester 2008), and these describe in detail the transport of Na^+ from the growing medium into the roots, the Na^+ loading into and unloading from the xylem, and its redistribution within the plant.

16.2.3 Plant Response to Oxidative Stress Induced by Salinity

In addition to its known components of osmotic stress and ion toxicity, salt stress is also manifested as an oxidative stress, all of which contribute to its deleterious effects (Hernandez et al. 2000; Mittova et al. 2002). Oxidative stress is characterised by the overproduction of reactive oxygen species (ROS) represented predominantly by superoxide anion, hydrogen peroxide, hydroxyl radical, and singlet oxygen. Plants have defensive mechanisms and utilise several biochemical strategies to avoid damage caused by ROS. Plant enzymatic defences include antioxidant enzymes such as the phenol peroxidase, ascorbate peroxidase, glutathione peroxidase, superoxide dismutase, and catalase which, together with other enzymes of the ascorbate-glutathione cycle, promote the scavenging of ROS (Hernandez et al. 2001; Hong et al. 2007). The biochemical defence system also includes carotenoids, ascorbate, glutathione and tocopherols. Several authors have suggested that the function of sugars, poliols, glycine-betaine and proline could be to protect cells against the hydroxyl radical (Sickler et al. 2007).

The sources of ROS under stress, mechanisms of ROS detoxification and the role of ROS in stress signalling are all active areas of current research and have been extensively studied and reviewed (Ben Amor et al. 2007). However, more studies are necessary before a definitive conclusion can be reached about the role of the ROS production under stress. Thus, increased ROS production has long been known under the heading of 'oxidative stress', which in itself is a negative term implying a harmful process. In contrast to this negative term, implying a state to be avoided, Foyer and Noctor (2005) proposed that the syndrome would be more usefully described as 'oxidative signalling', that is, an important and critical function associated with the mechanisms by which plant cells sense the environment and make appropriate adjustments to gene expression, metabolism and physiology.

16.2.4 Homeostasis and Protection or Damage Repair Induced by Salinity

Responses to salt stress are often discussed in terms of homeostasis and protection or damage repair (Zhu 2001). Mechanisms of ion homeostasis and osmotic homeostasis attempt to restore the cellular ion or water content to levels similar to those present under unstressed conditions. Protection and damage repair mechanisms attempt to prevent or repair cellular damage caused by altered ion or water content under stress. To understand the molecular mechanisms that plants have developed to cope with salinity, key genes for salt stress should be identified. Present engineering strategies for enhanced salt tolerance rely on the transfer of one or several genes either involved in signalling and regulatory pathways or encoding enzymes present in pathways leading to the synthesis of functional and structural protectants, such as osmolytes and antioxidants, as is indicated in the next sections. Osmotic homeostasis probably depends on the action of genes for solute synthesis, and a number of channels and carriers for uptake and compartmentalisation of inorganic solutes, especially K^+ (Rodriguez-Navarro 2000). Aquaporins may also have a role in osmotic homeostasis by facilitating water movement, but despite some progress achieved recently, not much is currently known about the levels of physiological relevant aquaporin function and regulation, and its effects on plant water balance (Kaldenhoff et al. 2008). Because of the most important effect induced by salinity to long-term is the toxic effect induced by the transport of saline ions to the shoot, most approaches have been directed to studying cation transporters and their regulation, especially the Na^+ transporters genes, like *SOS1* (Shi et al. 2002), *AtHKT1* (Rus et al. 2004) and *AtNHX1* (Apse and Blumwald 2002), whereas the Cl^- transport mechanisms in plants have been rarely studied (Colmenero-Flores et al. 2007).

In spite of the important work carried out on the Na^+ transport mechanism in plants in recent years, more advances are necessary to understand the role and regulation of genes involved in the re-establishment of Na^+ homeostasis under salt stress (Maathuis 2006; Pardo et al. 2006; Apse and Blumwald 2007; Munns and Tester 2008). Moreover, these genes should be identified not only in model species like *Arabidopsis* but also in crop species, since their role may be different depending on the species, as seems to occur in the *HKT1* genes. Recent studies on the role of *AtHKT1* in Na^+ transport have shown that this gene appears to control retrieval of Na^+ from the xylem before it reaches the shoot (Davenport et al. 2007). However, the *HKT* gene family is quite diverse, and this diversity led to early reports of apparently contradictory properties (see '*AtHKT1;1*, A case study of confusion', in Munns and Tester 2008). Increased clarity has been provided by dividing the *HKT* gene family into two distinct subfamilies (Platten et al. 2006). Recently, Nagata et al. (2008) did a comparative molecular-biological analysis of membrane transport genes in different organisms, ranging from bacteria to animals and plants. They compared the numbers of membrane transporter genes in *Arabidopsis* and rice. Although many transporter genes are similar in these plants, *Arabidopsis* has a more diverse array of genes for multi-efflux transport and for

response to stress signals, and rice has more secondary transporter genes for carbohydrate and nutrient transport.

Another of the genes controlling long-distance transport in *Arabidopsis thaliana* is the *SOS1* gene, which encodes a plasma membrane Na^+/H^+ antiporter that is essential for salt tolerance (Shi et al. 2003). Key intermediaries in the regulation of SOS1 are the protein kinase SOS2 and its associated Ca^{2+}-sensor protein SOS3 (Pardo et al. 2006). Co-expression of the three proteins in a yeast strain lacking endogenous Na^+ transporters restored salt tolerance to a much greater extent than SOS1 alone, whereas SOS2 or SOS3 individually failed to stimulate SOS1 activity (Quintero et al. 2002). Despite these advances, the function of *SOS1* in plants needs to be clarified, as this transporter seems to function in retrieving Na^+ from the xylem to prevent excess Na^+ accumulation in the shoot under severe salt stress (Shi et al. 2003). On the contrary, *SOS1* functions to load Na^+ into the xylem for controlled delivery to the shoot when salinity is moderate (Shi et al. 2002).

Na^+ has to be compartmentalised in the vacuole to prevent excess Na^+ accumulation in the cytoplasm. *Arabidopsis* AtNHX1 protein was the first Na^+/H^+ exchanger identified in plants (Gaxiola et al. 1999). The presence of Na^+/H^+ antiporter activities has been physiologically characterised in tonoplast vesicles and it is molecularly represented by six *Arabidopsis* genes, *AtNHX1–6* (Yokoi et al. 2002). *AtNHX1* steady-state transcript levels were increased in response to NaCl, KCl, sorbitol, and ABA, suggesting that *AtNHX1* transcript upregulation is not specific to ionic stress but is common to osmotic stress (Gaxiola et al. 1999; Shi and Zhu 2002). Overexpression of *NHX* antiporters has been used to improve salt tolerance in several plant species (Zhang and Blumwald 2001; Wu et al. 2004). Pardo et al. (2006) summarised the important work carried out in transformed plants with *NHX* antiporters and concluded that relevant information is still missing about the way *NHX* expression affects ion compartmentalisation within the cell, overall ion homeostasis, and osmotic adjustment, in order to fulfil the physiological premises basis that sustain the increase in salt tolerance.

With respect to the Na^+ pumps, there are no classical pumps in higher plants. In fungi and mosses there is a Na^+ pump, *ENA1*, which hydrolyses ATP to pump Na^+ out of the cell (Benito and Rodriguez-Navarro 2003). Also in yeast are the *HAL* genes (Serrano et al. 1999). Overexpression of *HAL1* conferred salt tolerance in yeast by facilitating intracellular K^+ accumulation and decreasing intracellular Na^+ (Rios et al. 1997). According to Munns (2005), the expression of these genes, which do not have orthologues in higher plants, may introduce new mechanisms for Na^+ and/or K^+ homeostasis. *HAL1* was introduced into tomato and its overexpression led to higher salt tolerance in the progeny of different transgenic plants, and furthermore, a similar mechanism to that in yeast was observed, namely by facilitating K^+/Na^+ selectivity under salt stress (Gisbert et al. 2000; Rus et al. 2001). Afterwards, a transgenic line with a very high gene expression level was selected in order to corroborate the tolerance induced by *HAL1* in tomato. However, the fruit yield of the homozygous plants was lower than that of the azygous plants, in spite of the lower Na^+ uptake and Na^+ translocation to the shoot that persisted over time in the homozygous line (Muñoz-Mayor et al. 2008).

Moreover, the ability of *HAL1* to reduce Na^+ uptake in the long-term was shown by the low accumulation of Na^+ in fruits. In-depth physiological characterisation of these plants showed that the greater ability for Na^+ exclusion in the homozygous line caused another type of osmotic problem, as leaves required increased synthesis of organic solutes to maintain osmotic balance, which has a high energy cost (Balibrea et al. 2003), and hence a growth penalty that negatively impacted on fruit yield. These results demonstrate the importance of considering the osmotic component of salt stress, especially when overexpression is used as a tool to identify genes involved in salt tolerance. It may be concluded that the constitutive expression of a gene may induce improvements in a trait, e.g. Na^+ exclusion, but this positive effect may provoke other physiological problems in the plant. In this sense, studies with *AtNHX1*-overexpressing tomato showed greater K^+ uptake than control lines but nevertheless were prone to K^+ deficiency symptoms at leaf K^+ concentrations greater than the control line (Leidi et al. 2005). According to Pardo et al. (2006), this paradoxical phenotype is likely to be due to exacerbated activity of the *AtNHX1* antiporter in the transgenic lines which could increase the vacuolar pool at the expense of the cytoplasmic K^+ pool, thereby inducing a K^+ starvation signal and eliciting greater K^+ uptake by roots. Thus, proper modulation of gene expression in time and space may be more important than mere overexpression of the transgene (Tonsor et al. 2005), as is pointed out below.

Taken together, tolerance to salinity stress is a complex phenomenon at both the cellular and the whole plant level, and a considerable gap still exists between the knowledge gained by physiological and molecular studies in response to salt stress and the knowledge required to develop crop plants with enhanced tolerance to field saline conditions. A focus on comprehensive physiological, molecular and metabolic aspects of salt stress in crop plants is needed to advance in the knowledge of the salt tolerance basis and to facilitate the development of crop plants with enhanced stress tolerance. New tools for functional genomics emerged in recent years may enhance significantly the descriptive power of physiological analysis (Sanchez et al. 2008; Weckwertha 2008), although it is absolutely necessary to elucidate not only processes involved in ionic stress tolerance and compatible solute synthesis, but also other processes as mechanisms of osmotic tolerance, which remain unknown (Munns and Tester 2008). Furthermore, to benefit more from the new genomic approaches, molecular studies with plants grown in physiologically realistic conditions are needed. Finally, it should be mentioned that differences in salt tolerance mechanisms between salt-sensitive glycophytes and salt-tolerant halophytes may result from changes in regulation of the same basic set of genes involved in salt tolerance. Many genes encoding potential salt tolerance determinants have been identified in model plant species, but a comparative study of the expression of those genes in both halophytes and glycophytes has been hampered by the lack of genetic and molecular tools available for the halophytic species. An important step forward has been provided by the use of *Arabidopsis* molecular and genetic tools to characterise the halophyte *T. halophila*, which is allowing have shown that the two species exhibit both shared and

divergent responses to salt stress (Gong et al. 2005a; Kant et al. 2006). Currently, we are trying to identify genes involved in salt tolerance in the wild salt-tolerant tomato species *S. pennellii*, using insertional mutagenesis as a genetic tool, as discussed in next sections.

16.3 Breeding for Salinity Tolerance

The first question that arises when breeding for salinity tolerance is the possibility to develop genotypes tolerant to salinity. Natural evolution has shaped flowering plant species and ecotypes to live under saline conditions, so that there is some compatibility between plant life and saline conditions and, therefore, it should be possible to artificially reproduce this process. However, the term 'tolerant genotype' depends upon the context. In natural environments a tolerant genotype is one that competes in the ecosystem and produce offspring. In an agricultural setting a tolerant genotype has to produce an economic yield (Yeo 2007). Breeders try to obtain genotypes not only able to live under saline conditions (biological tolerance) but also to grow and maintain agricultural productivity (agricultural tolerance) with the aid of appropriate cultivation methods. The success achieved in producing salt-tolerant varieties of crops has, however, been very limited. About 29 cultivars in only 12 species had been released for their salt tolerance until year 2000 (Flowers et al. 2000) and there has been little advance since then.

Obtaining salt-tolerant cultivars comprises three steps: (1) the presence of genetic variability for salinity tolerance in the species to be bred, in species that can be crossed with the target species, or even in organisms genetically far from the target species (2) to find out the gene or genes underlying the tolerance, and (3) the transmission of the gene(s) responsible of the tolerance to the cultivars.

16.3.1 Variability in Salinity Tolerance

All the crops where genetic variability for salt tolerance has been investigated have shown some degree of tolerance, either in the cultivated species or in closely related species that can be crossed with the cultivated one. Examples of species showing intra-specific variability are sorghum (Igartua and Garcia 1999), strawberry clover (Rumbaugh et al. 1993), rice (Alia et al. 2006), cotton (Ashraf 2002), lentil (Ashraf and Waheed 1990), etc. Inter-specific variability among genetically related species has been demonstrated in crops as different as tomato (Bolarin et al. 1991), durum wheat (Munns et al. 2000), eucalyptus (Niknam and McComb 2000) and many others.

Tolerance to salinity in higher plants affects numerous plant processes at all levels of organization (ion transport, osmotic adjustment, ion selectivity, nutrition, compartmentation, growth, water use, water use efficiency, etc.). Variability for osmotic adjustment has been demonstrated in wheat (Morgan 1992) barley (Blum 1989), rice

(Lilley and Ludlow 1996), sorghum (Basnayake et al. 1993), maize (Bolaños et al. 1993), tomato (Borsani et al. 2002) etc.; for Na^+ transport, in rice (Yeo 1992) and tomato (Cuartero et al. 2002). Cuartero et al. (2002) also demonstrated phenotypic variability in tomato for the relation [Na^+]/leaf area reduction, for the relation [K^+]/[Na^+] in leaf, and for selective Na^+ accumulation in old leaves.

It seems that the necessary variability for starting a breeding program is not hindering the development of salt tolerant cultivars in most crops. However "tolerance to salinity" is a very inaccurate term because it depends on the stage of plant development and where the tolerance has been measured. The easiest stage of development for determining tolerance to salinity is germination because experiments can be performed in the lab under controlled environmental conditions and with a high number of genotypes at the same time. But tolerance to salinity at germination stage has proved without relation to tolerance at adult stage in sorghum (Krishnamurthy et al. 2007), lentil (Ashraf and Waheed 1990), tomato (Foolad and Lin 1997), and rice (Moradi et al. 2003). Lack of relation between tolerance in vegetative stages and harvest has also been demonstrated (Greenway and Munns 1980; Shannon et al. 1987; Caro et al. 1991). Tolerance to salinity in one stage of development seems quite independent from tolerance to salinity in other stages what complicates indirect selection and comparison of results coming from different experiments and researchers.

Quantification of salt-tolerance is also difficult in the field, because (1) the stress may be experimentally uncontrollable because of variable rainfall, (2) the notoriously high field heterogeneity for salinity is liable to confound any planting plan designed for field experiments, and (3) salt uptake and sensitivity are modulated by environmental conditions that may affect each variety differently: any parameter which affects the transpiration rate (such as light intensity, temperature and humidity), can change dramatically a plant's susceptibility to salinity (Yeo et al. 1990). To avoid field problems measuring tolerance, plants are usually grown on inert substrates supplemented with a nutrient solution salinized with NaCl or a known mixture of salts enriched in NaCl.

Salt concentrations at which plants are grown to measure tolerance to salinity deserve also attention. The degree of salt sensitivity of genotypes at the germination stage is generally maintained when different salt concentrations are tested (Cuartero and Fernandez-Muñoz 1999; Foolad 2004), but this behaviour is not general and Cruz (1990) points out that the most tolerant tomato cultivars at 5–7 dS m^{-1} do not correspond with the most tolerant ones at 13 dS m^{-1}. It is necessary to define salinity conditions on the field and selection and experiments should be performed at those established salinity conditions (Cuartero et al. 2008).

16.3.2 Determinants Underlying Salinity Tolerance

It is known that most of the processes empirically determined to be important in plant tolerance to salinity exhibit quantitative inheritance (they show continuous variation) and a high degree of environmental influence (Cuartero and Fernandez-Muñoz 1999; Zhang et al. 1999; Mikiko et al. 2001).

Direct selection under field conditions for quantitative traits is difficult because fluctuating environmental factors adversely affect the accuracy and repeatability of such traits. Indirect selection for markers linked to quantitative traits has been suggested and put into practice for some time (Sax 1923; Falconer 1981). For an indirect selection marker to be useful in a breeding program, it has to exhibit both a significant genetic correlation with the trait and higher heritability than the trait itself (Falconer 1981). Molecular (DNA) and biochemical markers closely linked to quantitative trait loci (QTLs) affecting salt tolerance are good candidates because their expression are almost independent from the environment, and powerful biometric methods have been developed and applied to QTL mapping.

The main goal of QTL detection is marker-assisted selection (MAS) and QTL cloning (Asins 2002). The success in both of them depends on our ability to locate the QTL in chromosome regions as narrow as possible.

Salinity affects most of the biological processes in plants; hence tolerance to salinity is determined by many components. Some of them have been reviewed in the first part of this chapter, that is: accumulation of solutes (inorganic and organic) to maintain cell volume and turgor to counteract osmotic stress; ion exclusion from the root and ion compartmentalization to fight against ionic stress; preferential transport of water, K^+ and Ca^{2+} to restore nutritional status; and over-production of antioxidant enzymes and antioxidant compounds to avoid oxidative stress. All of those characters are quantitatively inherited and will require the detection of the QTL governing their expression.

The basis of QTL detection is the identification of significant statistical association between phenotypes and specific genetic markers, established by the joint analysis of segregation of markers and phenotypes in individuals or lines. The most extensively segregating generation used has been the F_2 because it can be rapidly obtained. Additive and dominant effects can be properly detected in F2 generations, but F_2 have two important drawbacks: the genotypes cannot be replicated and evaluated several times and in several environments, and epistatic interactions are frequently undetected. According to Asins (2002) recombinant inbred lines (RIL) or double haploids (DH) can be reproduced independently and continuously evaluated with respect to additional quantitative traits and markers, with all the information being cumulative. Additive and epistatic interaction can be determined with RIL and DH but not dominant or over-dominance effects.

When dealing with a complex character as tolerance to salinity that depends on a number of quantitative traits, QTL analysis allows an integrative approach. Total phenotypic variation can be explained in three main ways: the identification of QTLs partially controlling the tolerance by analysing multiple traits in the same segregating population; the contribution of QTL × Environment (E) interaction; and the contribution of QTL × QTL interactions or epistastatic effects. Such analyses under different salinity levels and plant stage(s) can only be properly and efficiently carried out using populations of DHs or RILs (Cuartero et al. 2006).

16.3.2.1 QTL × E Interaction

Since salt concentration in the soil is highly variable, it is frequent to test plant tolerance to salinity in several salt concentrations applied to the root system. Genotype × salt treatment interaction has been found in several occasions and species (Bolarin et al. 1991; Asins et al. 1993; Igartua 1995; Lee et al. 2004). It is then not surprising that QTL also show interaction with the environment, which means that expression of particular chromosome regions differs across environments and markers identified would be significant in only one or some of the environments. However, genotype × salt interactions are detected with ANOVA in the numerous experiments including several genotypes and two or more environmental conditions, but to detect QTL × E interaction it is necessary to study segregant populations with several replicates of each genotype in two or more environments, and unfortunately this kind of studies are rare. In addition F_2 population, the most frequent segregant population analyzed, is a poor plant material to detect QTL × E interactions and studies with RIL or DH populations in two or more environments are uncommon.

When appropriate segregant populations (RIL or DH) are grown in at least two salinity conditions (control and saline), more QTL have been identified in saline than in control conditions, and significant QTL × E interaction has been found in all the experiments designed to detect the interaction. Monforte et al. (1997) found only 6 QTL with expression in saline and control conditions, while 9 where expressed in the control and 18 specific for saline conditions. Takehisa et al. (2004) described 37 and 20 QTL in the fields irrigated with saline and fresh water respectively. Manneh et al. (2007) reported 8 markers expressed in control and saline conditions while 28 where expressed in only one of the conditions tested. Villalta et al. (2008) detected 4 QTL in control and 20 in saline conditions in one population, and 11 in saline, 2 in control and 2 in saline and control conditions in another population. Those examples have used only two experimental environments: control and saline conditions. It is necessary to asses that QTL detected under saline condition are expressed in different salt concentrations, otherwise QTL should be found for each specific salt concentration on which tolerant genotypes were to be grown.

QTL × E interaction, as genotype × environment interaction, has an evolutionary base because, according to Asins (2002), this sensitivity to the environment results in phenotypic plasticity (the possibility to take alternative development fates depending on environment) which is likely to be of particular importance in plants because they cannot move from one environment to another. But, from the viewpoint of the breeder, QTL × E interaction complicates the use of QTL in MAS for tolerance to salinity because salinity in the soils is not constant but variable, spatially in the field and also temporarily due to rain and irrigation variation from one year to another. To deal with this unpredictable situation breeders have to manage as many QTL related to tolerance to salinity as possible in order to breed salt tolerant genotypes. Some authors see the QTL × E interaction as lack of consistency of

QTL effects in different environments and suggest using only QTL expressed in most of the environments in MAS programmes. However, as mentioned before, there are many more QTL related to tolerance to salinity expressed only under saline conditions and, according to their importance, they are key QTL to breed salt tolerant genotypes.

The number of QTL related to some characters important in salt tolerance is high: 33 in the case of Monforte et al. (1997), 57 detected by Takehisa et al. (2004), 38 by Manneh et al. (2007), and 17 or 21, depending on the population considered, by Villalta et al. (2008). This number will probably increase when more characters and environmental conditions (salt concentrations) are considered, making very difficult to handle all of them in a MAS programme. Fortunately some QTL associations have been found when they have been mapped (Villalta et al. 2008) which will assist in the introduction of those QTL clusters in elite lines.

16.3.2.2 QTL × QTL Interaction

QTL detection experiments are notoriously poor at detecting interactions between loci (Frankel and Schork 1996). For this reason, we do not have a good idea as to how common such epistatic effects are, but when experiments have been designed to test specifically for their presence, these interactions have been found (Flint and Mott 2001). The molecular dissection of variation in bristle number in *Drosophila* indicated that the combination of two QTL had much larger effect than predicted from their individual effects (Gurganus et al. 1999). Epistasis has also been documented in plants. Lark et al. (1995) demonstrated epistasis in soybean, Eshed and Zamir (1996) in tomato, Lukens and Doebley (1999) in maize, Maheswaran et al. (2000) in rice, etc. The detection of epistasis seems to be independent of the species and the lack of information about QTL × QTL interaction could be explained by the plant material employed in the experiments. According to Asins (2002), epistasis between QTL can hardly be detected in advanced backcrosses or F_2 populations. The reason is that every backcross generation reduces the number of gene combinations while increasing genes from the recurrent genotype and in F_2 populations, even if large, there are few individuals with two-locus double homozygotes. RIL or DH populations are, definitely, the appropriate plant material to detect epistasis.

MAS will lose efficiency if QTL × QTL interactions do exist and are ignored in the selection process because they have not been detected and quantified.

16.3.3 Transmission of Determinants Responsible of Salt Tolerance

The traits governing tolerance to salinity are quantitatively inherited with low heritability (Cuartero et al. 2006). In addition, these traits are difficult to measure in segregating populations, requiring meticulous control of environmental variables

and replication of progenies over locations and seasons. MAS is an attractive approach when a few QTL control a significant portion of the variability for the traits under selection (Zhang et al. 1999).

Translating results from gene or QTL discovery to farm applications requires an agronomic approach rather than a purely academic perspective. We need not only an understanding of the genetic mechanisms underlying a particular trait but also a keen sense of the target environments and the appropriate germplasm to use as vehicles to deliver the traits (Leung 2008). It is important to note that cultivars bred for salt tolerance have to be not only salt tolerant, but also achieve the same desirable traits of productivity, quality, resistance to diseases, and adaptation to cultural techniques the current cultivars have. Elite lines, cultivars or parents of today's hybrids would be the ideal candidates to be improved for tolerance to salinity by introducing the tolerant QTL from the donors.

Today cultivars are the result of many years of selection and incorporation of traits demanded by consumers and growers. The delicate equilibrium among productivity, quality, resistance to diseases, and other agronomic traits showed by those cultivars could be probably altered with the introduction of a substantial number of QTL as those required for tolerance to salinity. Even with the help of codominant molecular markers tightly linked to the QTL to be transmitted, MAS programmes will need the phenotypic selection of the breeder generation after generation. Phenotypic selection can only be properly practised in generations approaching homocigosity slowly. Accordingly, MAS programmes should be designed to slowly approach the recurrent parent genotype. Cuartero and Fernández-Muñoz (1999) described an example of such a programme. A partial solution would be to breed rootstocks tolerant to salinity, grafting onto them the shoot of today cultivars. Martinez-Rodriguez et al. (2008) have recently demonstrated the beneficial effects of some rootstocks chosen because of their tolerance to salinity on tomato yield irrigated with saline water.

Despite innovations like better marker systems and improved genetic mapping strategies, the success of MAS has been very limited even taking into account that tolerance of the breeding lines is not expected to be as high as that of the tolerant donors. Manneh et al. (2007) pointed out the reliability of MAS to identify superior yielding rice genotypes under stress in the field. Several rice lines have been bred and released in the Philippines, Bangladesh, and India demonstrating significant yield advantages over salt-sensitive varieties (Ismail et al. 2007). Those two examples of success in breeding for salinity tolerance have been performed with rice. Rice has two important advantages over other species, especially dicotyledonous species: (1) tolerance to salinity is controlled by a few QTL with large effects, and (2) it is especially sensitive to salinity only during early seedling and reproduction stages (Leung 2008). The partial success on breeding salt tolerant rice varieties, a feasible plant model, should encourage the work with other species.

Salinity increases steadily in agricultural soils because of the irrigation water employed and because inadequate irrigation practices. If a soil salinity target has been fixed at the beginning of a breeding programme, it can substantially increase in the about 10 year's period needed to obtain a new cultivar with tolerance to

salinity. So, breeding salinity tolerance cultivars cannot be alone a long term solution to grow in saline soils but it must be complemented by other cultural techniques to stabilize the soil salinity level (Cuartero et al. 2008).

16.4 Improving Salt Tolerance Trough Gene Transformation

Despite great efforts to increase the level of salinity tolerance in species of agronomic interest, the results obtained through conventional breeding methods, as well as by some biotechnological approaches (e.g. *in vitro* selection) have been rather scarce (Flowers 2004; Yamaguchi and Blumwald 2005; Ashraf et al. 2008). The difficulty in obtaining practical results through these and other approaches explains the expectations generated to obtain salt tolerant cultivars via genetic transformation.

In numerous papers published from early 1990s until the present time, several authors have claimed enhancement of salt tolerance through either overexpression of endogenous genes or, more frequently, heterologous expression of genes that supposedly act on different mechanisms involved in the process (Cuartero et al. 2006).

Genes that have proven somewhat effective in providing stress tolerance using a transgenic approach belong to different categories. Preliminary research in this field was mainly focused on the overproduction of metabolically compatible (organic) solutes in transgenic plants (Chen and Murata 2002; Penna 2003; Kavi Kishor et al. 2005). More recently, attention has been paid to the modification of the glycine betaine biosynthesis pathway by using genes isolated from different sources (Su et al. 2006; Shirasawa et al. 2006; Zhou et al. 2007; Waditee et al. 2007; Park et al. 2007), most probably because the quaternary ammonium compound glycine betaine is accumulated in numerous halophytes from several families (Rhodes and Hanson 1993; Flowers and Colmer 2008). In this respect, it has been shown that the accumulation of glycine betaine in genetically modified plants of tomato is more effective in the chloroplasts than in the cytosol (Park et al. 2007), in a similar way to that previously observed in rice (Sakamoto et al. 1998). Notably, the accumulation of glycine betaine in transplastomic plants of carrot led to high levels of salt tolerance (up to 400 mM NaCl; Kumar et al. 2004).

Another strategy to increase the level of salt tolerance has been the transfer of genes codifying different kinds of proteins functionally related to macromolecules protection (LEA proteins, osmotin, chaperons, mRNA binding proteins; Wang et al. 2003; Zhang et al. 2007) or the protection of key metabolic enzymes (Arrillaga et al. 1998).

The scavenging of reactive oxygen intermediates through the transfer and expression of genes encoding detoxification enzymes (e.g. glutathione S-transferase, superoxide dismutase, ascorbate peroxidase, catalase) is an alternative way to protect cells against oxidative stress, thus limiting the damage produced by salt treatments (Apel and Hirt 2004). Interestingly the co-expression of more than one gene involved in oxidative stress protection in both the chloroplasts and cytosol gave rise to plants with increased tolerance to different types of abiotic stress (Zhao and Zang 2006; Tseng et al. 2007; Lee et al. 2007).

Genetic manipulation with genes encoding membrane proteins involved in the uptake and transport of water and ions, such as water channel proteins and ion transporters, is an alternative approach (Yamaguchi and Blumwald 2005; Chinnusamy et al. 2005; Forrest and Bhave 2007). As ion transport across the tonoplast into vacuoles is energised by a proton moving force (Gaxiola et al. 2007), the strategy based on the use of antiporters has generated high expectations in recent years. By overexpressing the vacuolar Na^+/H^+ antiport from *Arabidopsis thaliana* (*AtNHX1*), a high level of salt tolerance was reported in genetically modified plants of arabidopsis (Apse et al. 1999), tomato (Zhang and Blumwald 2001) and canola (Zhang et al. 2001). Thereafter, the *AtNHX1* gene was overexpressed in genetically modified plants of wheat (Xue et al. 2004), corn (Yin et al. 2004), beet (Yang et al. 2005), cotton (He et al. 2005) and tall fescue (Tian et al. 2006; Zhao et al. 2007), and in all the cases the authors indicated enhancement of salt tolerance. The greater tolerance conferred by the *AtNHX1* gene has been attributed to a process of Na^+ compartmentalisation into the vacuoles (Yamaguchi and Blumwald 2005). Despite being an attractive hypothesis, additional compelling evidence is needed before drawing a definitive conclusion. Since monovalent ions are judged toxic at the concentrations required for osmotic adjustment, it is generally assumed that Na^+ and Cl^- are compartmentalised in halophytes, predominantly in vacuoles, so that concentrations in the cytoplasm are maintained within tolerable limits. However, as stated by Flowers and Colmer (2008), experimental evidence for compartmentalisation of Na^+ into vacuoles is still limited, even in halophytes.

New *AtNHX* genes have been cloned and characterised (Yokoi et al. 2002; Aharon et al. 2003; Wang et al. 2007; Liu et al. 2008) and much effort has been made to identify orthologous genes in different species and perform the functional analyses, usually by overexpression in genetically modified plants. Examples are *AgNHX1* from *Atriplex gmelini* (Ohta et al. 2002); *BnNHX1* from *Brassica napus* (Wang et al. 2004); *GhNHX1* from *Gossypium hirsutum* (Wu et al. 2004); *OsNHX1* from *Oryza sativa* (Fukuda et al. 2004; Wu et al. 2005; Chen et al. 2007); *HbNHX1* from *Hordeum brevisubulatum* (Lu et al. 2005); *GmNHX1* from *Glycine max* (Sun et al. 2006; Li et al. 2006); *SsNHX1* from *Suaeda salsa* (Zhao et al. 2006b, c; Li et al. 2007); *PgNHX1* from *Pennisetum glaucum* (Verma et al. 2007; Rajagopal et al. 2007); *TNHX1* from *Triticum aestivum* (Brini et al. 2007); and *AeNHX1* from *Agropirum elongatum* (Qiao et al. 2007).

Overexpression of a vacuolar H^+-pyrophosphatase (*AVP1*) from *Arabidopsis thaliana* in transgenic plants of the same species increases the level of salt tolerance (Gaxiola et al. 2001). Similar results have been achieved by overexpressing the homologues from *Thellungiella halophila* (*TsVP*) in tobacco (Gao et al. 2006) and cotton (Lv et al. 2008), *Suaeda salsa* (*SsVP*) in arabidopsis (Guo et al. 2006), and *Triticum aestivum* (*TVP1*) also in arabidopsis (Brini et al. 2007). Interestingly, *TsVP* and *SsVP* genes have been cloned from halophytes (*Thellungiella halophila* and *Suaeda salsa*, respectively).

Likewise, a higher level of salt tolerance has been described through the overexpression of genes that codify plasma membrane Na^+/H^+ antiports cloned from different sources. For example, *AtSOS1* from *Arabidopsis thaliana* (Shi et al. 2003);

SOD2 from *Schizosaccharomyces pombe* (Gao et al. 2003; Zhao et al. 2006a); *nhaA* from *Escherichia coli* (Wu et al. 2005); and *OsSOS1* from *Oryza sativa* (Martínez-Atienza et al. 2007).

Regulatory genes such as transcription factors and those codifying signal transduction components or receptor-related proteins are another target (Kaur and Gupta 2005; Agarwal et al. 2006). Cloning of genes codifying transcription factors is a promising field as they lie upstream of many other genes. Recent research has allowed the identification of several transcription factors that are important in regulating plant stress responses, including not only different kinds of abiotic stress but also pathogen-induced defence responses, various physiological processes, hormonal signalling pathways and several developmental processes (Agarwal et al. 2006; Ham et al. 2006; Sohn et al. 2006; Seong et al. 2007; Ogawa et al. 2007; Liu et al. 2007; Dai et al. 2007; Nakashima et al. 2007).

It has also been suggested that genes codifying calcium sensors (Cheong et al. 2003) or even DNA helicases (pea DNA helicase 45, PDH45, Sanan-Mishra et al. 2005) and RNA helicases (DEAD-box helicase, Gong et al. 2005b; Owttrim 2006) could be involved in the process of salt tolerance. The role of siRNAs in stress conditions is also under study (Sunkar and Zhu 2004; Borsani et al. 2005; Jung and Kang 2007). Finally, knowledge of the processes related to DNA/RNA metabolism and G-protein signalling pathways could be useful in elucidating the less known stress signalling networks and thereby be helpful for engineering salinity-tolerance in crop plants (Tuteja 2007).

16.5 Functional Analysis of Salt Tolerance-Related Genes

Overall, the results obtained in this field show that the expression of different kinds of genes in transgenic plants can increase salinity tolerance, at least to some extent. Unfortunately, it is not possible to conclude for the moment that true halotolerant cultivars (i.e. with a sufficient tolerance level from an agronomic point of view) have been obtained via transformation. In fact, it would be best to avoid excessive optimism when drawing conclusions on the current state of this topic (Flowers 2004). Furthermore, when performing the functional analysis of a salt tolerance-related gene it would be advisable to take into consideration aspects such as the species used in the transformation, the procedure for evaluating the tolerance to salinity, and the complexity of the trait.

With respect to the first issue, the majority of transformation experiments have been carried out with the model species arabidopsis and tobacco (Vij and Tyagi 2007) which means we should be cautious when drawing conclusions. It would be best to perform such experiments in crops (Grover et al. 2003; Yamaguchi and Blumwald 2005; Cuartero et al. 2006; Bhatnagar-Mathur et al. 2008), as it is not certain what will occur when the genes that have given a positive result in models are expressed in cultivated species. The suitability of tobacco as a model in this field has been seriously questioned (Murthy and Tester 1996) and results from the evaluation of salt tolerance in transgenic arabidopsis plants cannot easily be

extrapolated to a crop species, as the important trait in the latter is the maintenance of production under stress. Extrapolations between different crop species cannot even be made, since the effects of salinity can be very different between species. If true advances are sought, the best approach is to focus efforts on cultivated species where the transformation technology is already available. Without doubt, the difficulties will be greater and the advances slower in crops than in model species, but the results will indicate the true importance of a certain transgene in the genetic context in which the tolerant phenotype will supposedly occur. Last but not least, any results could be of practical interest (Cuartero et al. 2006).

Regarding the procedure for evaluating the tolerance to salinity, if the published results are scrutinised, some of the methods of evaluation of transgenic materials appear of doubtful value. Results of a descriptive type or those based upon photographic evidence of the performance of plants may lead to confusing or erroneous conclusions (Flowers 2004). Responses to salt are frequently studied using small samples, in the very short-term, by using shock treatments and, furthermore, the data are collected during very specific growth periods (Yamaguchi and Blumwald 2005), in spite of the fact that in most crop species salt sensitivity depends on the growth stage (Perez-Alfocea et al. 1993; Khatum and Flowers 1995). Moreover, tolerance estimated on the basis of seed germination is not correlated with tolerance at later growth stages (Foolad and Lin 1997; Cuartero and Fernandez-Muñoz 1999). The usefulness of *in vitro* tests, frequently used for the evaluation of salt tolerance, could also be questioned because transpiring conditions have a major influence on Na^+ transport and tolerance (Moller and Tester 2007). However, a clear relationship between tolerance to salinity *in vitro* (callus) and *in vivo* (plants grown in greenhouse) has been observed for cultivated and wild tomato species (Perez-Alfocea et al. 1994; Cano et al. 1996) and similar results were obtained for cultivated species of *Cucumis* and *Citrullus* (melon, cucumber, and watermelon) and related wild species (Barage 2002). *In vitro* tests can provide complementary information on the effect of some transgenes (e.g. genes involved in ionic homeostasis) and can be useful for the pre-selection of transgenic lines (if an *in vitro* and *in vivo* correlation has previously been shown), but they should not be used as the only criterion to determine the degree of salt tolerance. In evaluating the tolerance of transgenic crops, it is important to perform long-term experiments, focus on growth and yield, and provide quantitative data (Flowers 2004; Munns and Tester 2008).

Bressan et al. (2008) have discussed the convenience of defining a minimum set of criteria for establishing unambiguously that transgenic plants do indeed show tolerance that is attributable to the transgene. In this respect, an important aspect in the functional analysis of a salt tolerance-related gene is the plant material to be used. The use of TG1 plants (primary transformants) is questionable because epigenetic effects (which are very important in some cases) may lead to erroneous conclusions. The evaluation in TG2 avoids the above problem, but it is necessary to take into account that this is a segregant progeny. In the authors' opinion, the best materials are the homozygous and azygous lines obtained in TG3. Thus, each homozygous line should be compared with two controls: the wild type and the corresponding azygous line without the transgene. Positional effects can generate

great differences in the expression of a given transgene in independent transgenic lines, indicating the necessity for selecting those with the best expression for the trait (Pineda 2005). Dose effects of the transgene can be estimated by comparing the behaviour of homozygous *versus* hemizygous lines (i.e. those derived from the sexual crossing between the homozygous and azygous lines). The relative tolerance of these lines can be estimated in the short- and mid-term, although, ultimately, the long-term response (estimating yield with quantitative data) must also be reported.

16.5.1 Complexity of the Trait and Sources of Genetic Variation

Salt tolerance is a complex trait (Bohnert et al. 1996; Tuteja 2007; Munns and Tester 2008). If one takes into account the diversity of mechanisms involved, the question that immediately arises is whether the introduction of a single gene can produce a sufficient level of tolerance or whether it is necessary to introduce several genes involved in different processes (e.g. osmotic adjustment, osmoprotection, ionic homeostasis, oxygen free radical scavengeing, stress response, restoration of enzymatic activity, photorespiration; Bohnert et al. 1996). Of course a particular gene (e.g. one that codes for a transcription factor) can have a cascade effect, modifying the expression of many genes. Alternatively, the expression of a gene involved in the compartmentalisation of ions in the vacuoles may alleviate toxic effects. Even so, it seems unlikely that a single gene could affect all the processes influenced by salinity. What is most likely is that the transference and expression, in a co-ordinated way, of a series of genes, each of which would affect one of the principal mechanisms of the process, would produce tolerant plants. The problem is that there is still not a clear idea of which genes have to be transferred.

When considering the future targets in this field, one can argue that rather than looking for salt tolerant-related genes in salt sensitive species, like arabidopsis, it would be better to focus on halotolerant plants. Flowers and Colmer (2008) have recently reviewed the mechanisms of tolerance in halophytes, plants that are able to survive and reproduce in environments where the salt concentration is around 200 mM NaCl or more. In this respect, the authors have proposed that research should be concentrated on a number of 'model' (halotolerant) species that are representative of the various mechanisms that might be involved in tolerance. As these halophytes are evolutionarily distant from the main crop species, from a breeding point of view it would perhaps be better to take advantage of the existence of halotolerant accesions of wild species related to a given crop, as occurs in the genus *Solanum* (Cuartero et al. 2006), *Citrullus* (Barage 2002), *Cucumis* (Barage et al. 2002) and many others. Unfortunately, despite the wealth of sources of variation, it is still not known which are the key genes determining the high level of salt tolerance in those plants.

16.6 Genomic Approaches for Dissection of Salinity Tolerance

The use of functional genomic approaches may serve to overcome the above mentioned problems. Transcriptomic analysis provides the expression profiles of hundreds or thousands of genes. At present, this kind of approach is being used to identify those genes that are expressed or deactivated in response to saline or other types of abiotic stress (Cuartero et al. 2006, and references therein). Although these methods might lead to an overestimation of the number of genes supposedly involved, which would make the identification of relevant genes among an enormous number of other genes with purely secondary or irrelevant functions more difficult, it is foreseeable that transcriptomic analysis will become a valuable tool in the near future. However, in order to fulfil the expectations created in this field, it would be sensible to take into account the stage of development at which the saline treatment is applied, to perform tissue-specific studies and to avoid traumatic or unnatural treatments (Munns 2005). In fact, transcriptomics studies can produce different answers depending on the tissue examined and whether the plant is growing or dying (Munns and Tester 2008). The relationship between the intensity of saline treatment and the degree of salt tolerance is another point to be considered. A high-salt treatment for a sensitive plant like arabidopsis will induce changes predominantly associated with senescence; however, a low-salt treatment may not result in discernable changes in gene expression (Munns and Tester 2008). The opposite situation should also be taken into account, as a high-salt treatment for a salt-tolerant plant (e.g. halophyte) may not produce any remarkable change in gene expression (as those plants grow normally in that situation and tolerate salt concentrations that kill or inhibit the growth of the majority of glycophytes) and a low-salt treatment may just reflect an abnormal situation for that species. Rather than apply these approaches either to model (salt-sensitive) species or halophytes, it would be better to apply them in both crop species and halotolerant accessions of related wild species and thus, by comparison, try to identify the genes responsible for tolerance (Bohnert et al. 2006; Cuartero et al. 2006).

Other genomic approaches should provide very useful information. For example, major advances have been achieved in the study of mechanisms of post-transcriptional gene silencing and high throughput systems are available to infer gene function (Baulcombe 2004; Herr et al. 2005; Cherian et al. 2006). We foresee that, if systematically applied in a large-scale program using halotolerant plants, this approach would be particularly valuable for the identification of genes involved in different mechanisms related to salt tolerance.

Overexpression has hitherto been the most widely used strategy for both the functional analysis of candidate genes as well as for the increase in salt tolerance in transgenic plants. The underlying idea is that by overexpressing a certain gene or by expressing it in a constitutive way it would always have a positive effect on the phenotype. But increasing evidence supports the idea that sometimes strong and constitutive promoters (e.g. CaMV-35S, mostly for dicots, or actin and ubiquitin1, for monocot species) involve a high energetic cost and yield a penalty in transgenic plants (Rus et al. 2001; Grover et al. 2003; Pineda 2005; Muñoz-Mayor et al. 2008) and, in other cases, the beneficial effects of the transgene are masked by pleiotropic

effects derived from the use of strong promoters (Romero et al. 1997; Capell et al. 1998; Kasuga et al. 1999). Yield penalty and/or pleiotropy not only make the interpretation of the results on the functional analysis of a given transgene difficult but also hamper any practical application. In fact, evidence from research in this field supports the advantages of using inducible promoters (Kasuga et al. 1999; Garg et al. 2002; Rohila et al. 2002; Lee et al. 2003; Su and Wu 2004; Nakashima et al. 2007). Moreover, the use of inducible or specific promoters will be essential when tackling the co-transference and co-expression of several genes to avoid homology-based gene silencing (Grover et al. 2003; Cuartero et al. 2006). It is expected that the identification of new *cis* regulatory elements, which allow proper expression in time and space, will be a major target in the near future (Yamaguchi and Blumwald 2005; Cherian et al. 2006; Cuartero et al. 2006; Bhatnagar-Mathur et al. 2008).

The use of mutants as genomic tools should also be one of the main research areas in the coming years. One of the key factors explaining our present knowledge in several areas of plant development lies in the detection and characterisation of mutants that have altered developmental traits, for example those affected in tomato fruit development and maturation (Emmanuel and Levy 2002; Tanksley 2004; Giovannoni 2007; Lozano et al. 2009). By comparison, the number of mutants with the level of salt tolerance affected in other species than arabidopsis which are already available for the scientific community is rather scarce, perhaps due to the difficulty in performing proper evaluation of the trait. Occasional spontaneous mutants or, alternatively, those generated by chemical (e.g. EMS) or physical (e.g. fast neutrons) methods could provide the basis for advancing the knowledge of physiological processes related to salt tolerance. However, in the absence of obvious candidate genes, the isolation of a gene altered in the mutant through a positional cloning strategy requires huge effort. Fortunately, at present we can overcome these problems by using alternative approaches.

Insertional mutagenesis with T-DNA or transposable elements constitutes a basic tool for the identification of genes and the analysis of their function. With respect to insertional mutagenesis with T-DNA, we can approach the tagging of genes by using a simple construction with a marker gene. In this way, the integration of T-DNA within the structural sequence or the controlling elements of a given gene will lead to its disruption and the consequent loss-of-function or, depending of the characteristics of the T-DNA-insert, gain-of-function or change in its level of expression (Krysan et al. 1999). Upon detecting the mutant phenotype in TG1 (in the case of dominant, semidominant or additive effects) or TG2 (recessive), its cloning can be easier as the gene is tagged by the T-DNA.

By comparison with classical insertional mutagenesis, trapping systems (Springer 2000) can be particularly useful for the identification of genes related to salt tolerance. The advantage of using enhancer, promoter or gene traps resides in its self dual nature. Like any other T-DNA, those traps act as insertion mutagens, but also, when T-DNA is integrated inside an endogenous gene in the appropriate orientation, the reporter gene lies under the control of the regulatory elements of the tagged gene. By analysing the reporter gene expression one can obtain a precise picture of the spatial and temporal expression pattern of the endogenous gene tagged by the trap. In this respect, trapping strategies bring great advantages over insertional mutagenesis

by allowing identification of functionally redundant genes, those expressed at multiple developmental stages (generating confusion during phenotyping), genes whose disruption causes early lethality, and genes whose disruption causes a soft phenotype that may not be detected (in this case, the reporter expression gives a clue to identify the phenotype during evaluation). In addition, gene identification is independent of the expression level of the gene, avoiding the risk of rejecting genes that are expressed at low levels, even though they have major effects on the phenotype. Finally, this is the best way to identify genes that are activated or repressed in response either to an external stimulus or biotic and abiotic stress situations.

Using an enhancer trap (kindly provided by Dr. Thomas Jack, Department of Biological Sciences, Dartmouth College, USA) and a promoter trap (developed in the laboratory of Drs. Rafael Lozano and Trinidad Angosto, Universidad de Almería, Spain), a collection of more than 2,000 T-DNA lines of halotolerant accessions of the wild tomato-related species *Solanum pennellii* has been generated (Anton et al 2009). Scrutiny of this collection to identify insertion mutants with altered levels of saline stress (mainly hypersensitive) is under way. Given that the collection of T-DNA lines is going to enlarge progressively, identification of new salt tolerance related-genes is expected in the near future. It is foreseeable that the use of these genomic approaches will allow the genetic dissection of the trait and, thereafter, the proper design of a breeding program.

16.7 Conclusions

A significant gap still exists between the knowledge gained in physiological and molecular studies of responses to salt stress, and the knowledge required to develop crop plants with enhanced tolerance to saline conditions. A focus on physiological, molecular and metabolic aspects of salt stress in crop plants would help to bridge this gap. It is still necessary to elucidate some processes involved in ionic stress tolerance, synthesis of compatible solute and also other processes as the mechanisms of osmotic tolerance. New tools arising from functional genomics may significantly enhance the descriptive power of physiological analysis and molecular studies with plants grown under realistic saline conditions.

It seems then that the necessary variability for starting breeding programs is not hindering the development of salt tolerant cultivars in most crops, although it is necessary to look for variability to some key physiological traits related to osmotic and ionic stresses, in specifically salt concentrations and plant development stages. QTL detection and mapping requires the development of large RIL or DH generations from parents with a wide variation in a number of traits related to tolerance to salinity. These populations will help to: (1) locate the QTL in chromosome regions as narrow as possible, (2) accumulate QTL information gained in different experiments and growing seasons, and (3) quantify QTL × E and QTL × QTL interactions. Elite cultivars or elite lines, parents of today's hybrids, are the ideal candidates to receive the salt tolerance QTL from donor genotypes. Introduction of QTL should be made by combining

marker assisted and phenotypic selection in breeding programs approaching slowly the recurrent elite lines. This will allow breeders to maintain the delicate equilibrium among productivity, quality, resistance to diseases, and other agronomic traits showed by elite cultivars and lines, that could be otherwise be altered with the introduction of a substantial number of QTL as those required to obtain tolerance to salinity.

Expression of some transgenes promotes a high level of salt tolerance in different species. Despite these interesting results, it is not possible to conclude yet that cultivars tolerant enough from an agronomic point of view have been obtained via transformation. To fulfil the expectations created by the plant transformation technique, it is necessary to advance on different aspects: (1) identification of genes actually involved in the process of salt tolerance, (2) isolation of genes from adequate sources of variation (wild species related to a given crop), (3) design vectors that allow the transfer and coordinate expression of several genes (since salt tolerance is a complex trait), and (4) identification of regulatory elements modulating spatially and temporarily the expression level of the transgenes. Despite present limitations, it is foreseeable that breeding programmes will benefit from ongoing functional genomics projects that could allow the use of genetic transformation as a regular breeding tool.

Acknowledgments The authors like to thank Dr. Eva Domínguez, Estación Experimental La Mayora, CSIC, for valuable suggestions while editing the manuscript. V. Moreno, B. Pineda and M.C. Bolarín would like to thank the CICYT (Ministerio de Ciencia e Innovación, Spain) for financial support (project AGL2006 15290). V. Moreno would also like to thank Dr. Thomas Jack, Department of Biological Sciences, Dartmouth College, USA for providing the pD991 vector.

References

Agarwal PK, Agarwal P, Reddy MK et al (2006) Role of DREB transcription factors in abiotic and biotic stress tolerance in plants. Plant Cell Rep 25:1263–1274

Aharon GS, Apse MP, Duan S et al (2003) Characterization of a family of vacuolar Na^+/H^+ antiporters in *Arabidopsis thaliana*. Plant Soil 253:245–256

Alarcon JJ, Sanchez-Blanco MJ, Bolarin MC et al (1993) Water relations and osmotic adjustment in *Lycopersicon esculentum* and *L. pennellii* during short-term salt exposure and recovery. Physiol Plant 89:441–447

Alia AJ, Xua JL, Ismail AM et al (2006) Hidden diversity for abiotic and biotic stress tolerances in the primary gene pool of rice revealed by a large backcross breeding program. Field Crops Res 97:66–76

Anton T, Garcia-Abellan JO, Perez F et al (2009) Identification of genes related to salt and drought tolerance in T-DNA tagging lines of cultivated and wild tomato species. In: Plant abiotic stress tolerance. Proc of Internat Conferen, Vienna feb 2009:161

Apel K, Hirt H (2004) Reactive oxygen species: metabolism, oxidative stress, and signal transduction. Annu Rev Plant Biol 55:373–399

Apse MP, Blumwald E (2002) Engineering salt tolerance in plants. Curr Opin Biotechnol 13:146–150

Apse MP, Blumwald E (2007) Na^+ transport in plants. FEBS Lett 581:2247–2254

Apse MP, Aharon GS, Snedden WA et al (1999) Salt tolerance conferred by overexpression of a vacuolar Na^+/H^+ antiport in Arabidopsis thaliana. Science 285:1256–1258

Arrillaga I, Gil-Mascarell R, Gisbert C et al (1998) Expression of the yeast HAL2 gene in tomato increases the in vitro ssalt tolerance of transgenic progenies. Plant Sci 136:219–226

Ashraf M (2002) Salt tolerance of cotton: some new advances. Crit Rev Plant Sci 21:1–30

Ashraf M, Waheed A (1990) Screening of local/exotic accessions of lentil (*Lens culinaris* Medic.) for salt tolerance at two growth stages. Plant Soil 128:167–176
Ashraf M, Athar HR, Harris RJC et al (2008) Some prospective strategies for improving crop salt tolerance. Adv Agron 97:45–110
Asins MJ (2002) Present and future of quantitative trait locus analysis in plant breeding. Plant Breed 121:281–291
Asins MJ, Bretó MP, Cambra M et al (1993) Salt tolerante in *Lycopersicon* species. II. Genetics effects and a search for associated traits. Theor Appl Genet 86:769–774
Balibrea ME, Cuartero J, Bolarin MC et al (2003) Sucrolytic activities during fruit development of *Lycopersicon* genotypes differing in tolerance to salinity. Physiol Plant 118:38–46
Barage M (2002) Identificación de fuentes de tolerancia a la salinidad y al estrés hídrico en especies silvestres de la familia *Cucurbitaceae*. Ph.D. thesis. Universidad Politécnica de Valencia, Valencia, Spain
Barage M, Pineda B, Atares A et al (2002) Identificación de fuentes de tolerancia a la salinidad en especies silvestres de la familia *Cucurbitaceae* sobre la base de parámetros de crecimiento vegetativo. Actas de Hrticultura 34:133–138
Bartels D, Dunkar R (2005) Drought and salt tolerance in plants. Crit Rev Plant Sci 24:23–58
Basnayake J, Ludlow MM, Cooper M et al (1993) Genotypic variation among adjustment and desiccation tolerance in contrasting sorghum inbred lines. Field Crops Res 35:51–62
Baulcombe D (2004) RNA silencing in plants. Nature 431:356–363
Ben Amor N, Jimenez A, Megdiche W et al (2007) Kinetics of antioxidant response to salinity in the halophyte *Cakile maritima*. J Integr Plant Biol 49:1–11
Benito B, Rodriguez-Navarro A (2003) Molecular cloning and characterization of a sodium-pump ATPase of the moss *Physcomitrella patens*. Plant J 36:382–389
Bhatnagar-Mathur P, Vadez V, Sharma KK (2008) Transgenic approaches for abiotic stress tolerance in plants: retrospect and prospects. Plant Cell Rep 27:411–424
Blum A (1989) Osmotic adjustment and growth of barley genotypes under drought stress. Crop Sci 29:230–233
Blum A, Munns R, Passioura JB et al (1996) Genetically engineered plants resistant to soil drying and salt stress: How to interpret osmotic relations? Plant Physiol 110:1051
Bohnert HJ, Gong Q, Li P et al (2006) Unraveling abiotic stress tolerance mechanisms - getting genomics going. Curr Opin Plant Biol 9:180–188
Bohnert HJ, Golldack D, Ishitani M et al (1996) Salt tolerance engineering requires multiple gene transfer. In: Collins GB, Shepherd RJ (eds) Engineering plants for commercial products and application. Ann New York Acad Sci 729:115–125
Bolaños J, Edmeades GO, Martinez L (1993) Eight cycles of selection for drought tolerance in lowland tropical maize. III. Responses in drought-adaptive physiological and morphological traits. Field Crops Res 31:269–286
Bolarin MC, Fernández FG, Cuartero J et al (1991) Salinity tolerance in four wild tomato species using vegetative yield-salinity response curves. J Am Soc Hort Sci 116:286–290
Borsani O, Cuartero J, Valpuesta V et al (2002) Tomato tos1 mutation identifies a gene essential for osmotic tolerance and abscisic acid sensitivity. Plant J 32:905–914
Borsani O, Zhu J, Verslues PE et al (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*. Cell 123:1279–1291
Bressan RA, Bohner HJ, Hasegawa PM (2008) Genetic engineering for salinity stres tolerance. Adv Plant Biochem Mol Biol 1:347–384
Brini F, Hanin M, Mezghani I et al (2007) Overexpression of wheat Na^+/H^+ antiporter TNHX1 and H^+-pyrophosphatase TVP1 improve salt- and drought-stress tolerance in *Arabidopsis thaliana* plants. J Exp Bot 58:301–308
Cano EA, Perez-Alfocea F, Moreno V et al (1996) Responses to NaCl stress of cultivated and wild species and their hybrids in callus culture. Plant Cell Rep 15:791–794
Capell T, Escobar C, Lui H et al (1998) Overexpression of the oat arginine decarboxylase cDNA in transgenic rice (*Oryza sativa* L.) affects normal development patterns in vitro and results in putrescine accumulation in transgenic plants. Theor Appl Genet 97:246–254

Caro M, Cruz V, Cuartero J et al (1991) Salinity tolerance of normal-fruited and cherry tomato cultivars. Plant Soil 136:249–255

Chen H, An R, Tang JH et al (2007) Over-expression of a vacuolar Na^+/H^+ antiporter gene improves salt tolerance in an upland rice. Mol Breed 19:215–225

Chen THH, Murata N (2002) Enhancement of tolerance of abiotic stress by metabolic engineering of betaines and other compatible solutes. Curr Opin Plant Biol 5:250–257

Cheong YH, Kim KN, Pandey GK et al (2003) CBL1, a calcium sensor that differentially regulates salt, drought, and cold responses in Arabidopsis. Plant Cell 15:1833–1845

Cherian S, Reddy MP, Ferreira RB (2006) Transgenic plants with improved dehydration-stress tolerance: progress and future prospects. Biol Plant 50:481–495

Chinnusamy V, Jagendorf A, Zhu JK (2005) Understanding and improving salt tolerance in plants. Crop Sci 45:437–448

Colmenero-Flores JM, Martinez G, Gamba G et al (2007) Identification and functional characterization of cation-chloride cotransporters in plants. Plant J 50:278–292

Cruz V (1990) Tolerancia a la salinidad y criterios de selección en *Lycopersicon* Mill. spp. Tesis doctoral. Universidad de Málaga, 484 p

Cuartero J, Fernández-Muñoz R (1999) Tomato and salinity. Sci Hort 78:83–125

Cuartero J, Bolarin MC, Asíns MJ et al (2006) Increasing salt tolerance in the tomato. J Exp Bot 57:1045–1058

Cuartero J, Bolarin MC, Moreno V et al (2008) Tolerancia a la salinidad. In: Moreno MT, Cubero JI, Atienza S, et al (eds) El ambiente y los estreses abióticos en mejora vegetal. IFAPA-Junta de Andalucía:231–262

Cuartero J, Romero-Aranda R, Yeo AR et al (2002) Variability for some physiological characters affecting salt tolerance in tomato. Acta Hort 573:435–441

Dai X, Xu Y, Ma Q et al (2007) Overexpression of an R1R2R3 MYB gene, OsMYB3R-2, increases tolerance to freezing, drought, and salt stress in transgenic *Arabidopsis*. Plant Physiol 143:1739–1751

Davenport RJ, Muñoz-Mayor A, Jha D et al (2007) The Na^+ transporter AtHKT1;1 controls retrieval of Na^+ from the xylem in *Arabidopsis*. Plant Cell Environ 30:497–507

De Costa W, Zorb C, Hartung W et al (2007) Salt resistance is determined by osmotic adjustment and abscisic acid in newly developed maize hybrids in the first phase of salt stress. Physiol Plant 131:311–321

Emmanuel E, Levy AA (2002) Tomato mutants as genomic tools. Curr Opin Plant Biol 5:112–117

Eshed Y, Zamir D (1996) Less-than-additive epistatic interactions of quantitative trait loci in tomato. Genetics 143:1807–1817

Estañ MT, Martinez-Rodriguez MM, Perez-Alfocea F et al (2005) Grafting raises the salt tolerance of tomato through limiting the transport of sodium and chloride to the shoot. J Exp Bot 56:703–712

Falconer DS (1981) Introduction to quantitative genetics. Longman, New York

FAO (2005) Global Network of Integrated Soil Management for Sustainability Use of Salt-Affected Soils. Rome FAO Land and Plant Nutrition Management Service http://www.fao.org/ag/agl/agll/spush

Flint J, Mott R (2001) Finding the molecular basis of quantitative traits: successes and pitfalls. Nat Genet Rev 2:437–445

Flowers TJ (1972a) Salt tolerance in *Suaeda maritima* L. Dum. The effect of sodium chloride on growth, respiration and soluble enzymes in a comparative study with Pisum sativum. J Exp Bot 23:310–321

Flowers TJ (1972b) The effect of sodium chloride on enzyme activities from four halophyte species of Chenopodiaceae. Phytochem 11:1881–1886

Flowers TJ (2004) Improving crop salt tolerance. J Exp Bot 55:307–319

Flowers TJ, Colmer TD (2008) Salinity tolerance in halophytes. New Phytol 179:945–963

Flowers TJ, Yeo AR (1995) Breeding for salinity tolerance in crop plants: Where next? Aust J Plant Physiol 22:875–884

Flowers TJ, Flowers SA, Hajibagheri MA et al (1991) Ion accumulation in the cell walls of rice plants growing under saline conditions: evidence for the Oertli hypothesis. Plant Cell Environ 14:319–325

Flowers TJ, Koyama ML, Flowers SA et al (2000) QTL: their place in engineering tolerance of rice to salinity. J Exp Bot 51:99–106
Foolad MR (2004) Recent advances in genetics of salt tolerance in tomato. Plant Cell Tissue Organ Cult 76:101–119
Foolad MR, Lin GY (1997) Absence of a genetic relationship between salt tolerance during seed germination and vegetative growth in tomato. Plant Breed 116:363–367
Forrest KL, Bhave M (2007) Major intrinsic proteins (MIPs) in plants: a complex gene family with major impacts on plant phenotype. Funct Integr Genomics 7:263–289
Foyer CH, Noctor G (2005) Oxidant and antioxidant signalling in plants: a re-evaluation of the concept of oxidative stress in a physiological context. Plant Cell Environ 28:1056–1071
Frankel WN, Schork NJ (1996) Who's afraid of epistasis? Nat Genet 14:371–373
Fukuda A, Nakamura A, Tagiri A et al (2004) Function, intracellular localization and the importance in salt tolerance of a vacuolar Na^+/H^+ antiporter from rice. Plant Cell Physiol 45:146–159
Gao F, Gao Q, Duan XG et al (2006) Cloning of an H^+-PPase gene from *Thellungiella halophila* and its heterologous expression to improve tobacco salt tolerance. J Exp Bot 57:3259–3270
Gao XH, Ren ZH, Zhao YX et al (2003) Overexpression of SOD2 increases salt tolerance of *Arabidopsis*. Plant Physiol 133:1873–1881
Garg AK, Kim JK, Owens TG et al (2002) Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses. Proc Natl Acad Sci USA 99:15898–15903
Gaxiola RA, Li JL, Undurraga S et al (2001) Drought and salt tolerant plants result from overexpression of the AVP1 H^+ pump. Proc Natl Acad Sci USA 98:11444–11449
Gaxiola RA, Palmgren MG, Schumacher K (2007) Plant proton pumps. FEBS Lett 581:2204–2214
Gaxiola RA, Rao R, Sherman A et al (1999) The Arabidopsis thaliana proton transporters, AtNhx1 and Avp1, can function in cation detoxification in yeast. PNAS-USA 96:1480–1485
Ge LF, Chao DY, Shi M et al (2008) Overexpression of the trehalose-6-phosphate phosphatase gene *OsTPP1* confers stress tolerance in rice and results in the activation of stress responsive genes. Planta 228:191–201
Giovannoni JJ (2007) Fruit ripening mutants yield insights into ripening control. Curr Opin Plant Biol 10:283–289
Gisbert C, Rus AM, Bolarin MC et al (2000) The yeast *HAL1* gene improves salt tolerance of transgenic tomato. Plant Physiol 123:393–402
Glenn EP, Brown JJ, Blumwald E (1999) Salt tolerance and crop potential of halophytes. Crit Rev Plant Sci 18:227–255
Gong Q, Li P, Ma S et al (2005a) Salinity stress adaptation competence in the extremophile *Thellungiella halophila* in comparison with its relative *Arabidopsis thaliana*. Plant J 44:826–839
Gong Z, Dong CH, Lee H et al (2005b) A DEAD box RNA helicase is essential form RNA export and important for development and stress responses in *Arabidopsis*. Plant Cell 17:256–267
Gorham J, Bristol A, Young EM et al (1990) Salt tolerance in the Triticeae: K/Na discrimination in barley. J Exp Bot 41:1095–1101
Greenway H, Munns R (1980) Mechanisms of salt tolerance in non-halophytes. Annu Rev Plant Physiol Plant Mol Biol 31:149–190
Grover A, Aggarwal PK, Kapoor A et al (2003) Addressing abiotic stresses in agriculture through transgenic technology. Curr Sci 84:355–367
Guo S, Yin H, Zhang X et al (2006) Molecular cloning and characterization of a vacuolar H^+-pyrophosphatase gene, SsVP, from the halophyte *Suaeda salsa* and its overexpression increases salt and drought tolerance of *Arabidopsis*. Plant Mol Biol 60:41–50
Gurganus MC, Nuzhdin SV, Leips JW et al (1999) High-resolution mapping of quantitative trait loci for sternopleural bristle number in *Drosophila melanogaster*. Genetics 152:1585–1604
Ham BK, Park JM, Lee SB et al (2006) Tobacco Tsip1, a DnaJ-Type Zn finger protein, is recruited to and potentiates Tsi1-mediated transcriptional activation. Plant Cell 18:2005–2020
He CX, Yan JQ, Shen GX et al (2005) Expression of an arabidopsis vacuolar sodium/proton antiporter gene in cotton improves photosynthetic performance under salt conditions and increases fibre yield in the field. Plant Cell Physiol 46:1848–1854
Hernandez JA, Jimenez A, Mullineaux P et al (2000) Tolerance of pea (*Pisum sativum* L.) to long-term salt stress is associated with induction of antioxidant defences. Plant Cell Environ 23:853–862

Hernandez JA, Ferrer MA, Jimenez A et al (2001) Antioxidant systems and O-2(.-)/H2O2 production in the apoplast of pea leaves. Its relation with salt-induced necrotic lesions in minor veins. Plant Physiol 127:817–831

Herr AJ, Jensen MB, Dalmay T et al (2005) RNA polymerase IV directs silencing of endogenous DNA. Science 308:118–120

Hong C, Hsu YT, Tsai YH et al (2007) Expression of ASCORBATE PEROXIDASE 8 in roots of rice (*Oryza sativa* L.) seedlings in response to NaCl. J Exp Bot 58:3273–3283

Igartua E (1995) Choice of selection environments for improving crops yields in saline areas. Theor Appl Genet 91:1061–1021

Igartua E, Gracia P (1999) Divergent selection for salinity tolerance at germination-emergence stage in grain sorghum. Maydica 43:161–168

Ismail AM, Heuer S, Thomson MJ et al (2007) Genetic and genomics approaches to develop rice germplasm for problem soils. Plant Mol Biol 65:547–550

Jung HJ, Kang H (2007) Expression and functional analyses of microRNA417 in *Arabidopsis thaliana* under stress conditions. Plant Physiol Biochem 45:805–811

Kaldenhoff R, Ribas-Carbo M, Sans JF et al (2008) Aquaporins and plant water balance. Plant Cell Environ 31:658–666

Kant S, Kant P, Raveh E et al (2006) Evidence that differential gene expression between the halophyte *Thellungiella halophila* and *Arabidopsis thaliana* is responsible for higher levels of the compatible osmolyte proline and tight control of Na^+ uptake in T. halophila. Plant Cell Environ 29:120–134

Kasuga M, Liu Q, Miura S et al (1999) Improving plant drought, salt and freezing tolerance by gene transfer of a single stress-inducible transcription factor. Nat Biotechnol 17:287–291

Kaur N, Gupta AK (2005) Signal transduction pathways under abiotic stresses in plants. Curr Sci 88:1771–1780

Kavi Kishor PB, Sangam S, Amrutha RN et al (2005) Regulation of proline biosynthesis, degradation, uptake and transport in higher plants: its implications in plant growth and abiotic stress tolerance. Curr Sci 88:424–438

Khatum S, Flowers TJ (1995) Effects of salinity on seed set in rice. Plant Cell Environ 18:61–67

Krishnamurthy L, Serraj R, Hash CT et al (2007) Screening sorghum genotypes for salinity tolerant biomass production. Euphytica 156:15–24

Krysan PJ, Young JC, Sussman MR (1999) T-DNA as an insertional mutagen in arabidopsis. Plant Cell 11:2283–2290

Kumar S, Dhingra A, Daniell H (2004) Plastid-expressed betaine aldehyde dehydrogenase gene in carrot cultured cells, roots, and leaves confers enhanced salt tolerance. Plant Physiol 136:2843–2854

Lark KG, Chase K, Adler F et al (1995) Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. Proc Natl Acad Sci USA 92:4656–4660

Lee GJ, Boerma HR, Villagarcia MR et al (2004) A major QTL conditioning salt tolerance in S-100 soybean and descendent cultivars. Theor Appl Genet 109:1610–1619

Lee JT, Prasad V, Yang PT et al (2003) Expression of Arabidopsis CBF1 regulated by an ABA/stress inducible promoter in transgenic tomato confers stress tolerance without affecting yield. Plant Cell Environ 26:1181–1190

Lee YP, Kim SH, Bang JW et al (2007) Enhanced tolerance to oxidative stress in transgenic tobacco plants expressing three antioxidant enzymes in chloroplasts. Plant Cell Rep 26:591–598

Leidi EO, Barragán V, Cubero B et al (2005) Function of endosomal NHX antiporters on plant K nutrition. Society for Experimental Biology Annual Main Meeting. Comp Biochem Physiol-Part A 141:S342

Leung H (2008) Stressed genomics - bringing relief to rice fields. Curr Opin Plant Biol 11:201–208

Lewis IN (1984) A vital resource in danger. Calif Agric 38:2

Li J, Jiang G, Huang P et al (2007) Overexpression of the Na^+/H^+ antiporter gene from *Suaeda salsa* confers cold and salt tolerance to transgenic *Arabidopsis thaliana*. Plant Cell Tissue Organ Cult 90:41–48

Li WYF, Wong FL, Tsai SN et al (2006) Tonoplast-located GmCLC1 and GmNHX1 from soybean enhance NaCl tolerance in transgenic bright yellow (BY)-2 cells. Plant Cell Environ 29:1122–1137

Lilley JM, Ludlow MM (1996) Expression of osmotic adjustment and dehydration tolerance in diverse rice lines. Field Crops Res 48:185–197

Liu H, Wang QQ, Yu MM et al (2008) Transgenic salt-tolerant sugar beet (*Beta vulgaris* L.) constitutively expressing an Arabidopsis thaliana vacuolar Na^+/H^+ antiporter gene, AtNHX3, accumulates more soluble sugar but less salt in storage roots. Plant Cell Environ 31:1325–1334

Liu K, Wang L, Xu Y et al (2007) Overexpression of OsCOIN, a putative cold inducible zinc finger protein, increased tolerance to chilling, salt and drought, and enhanced proline level in rice. Planta 226:1007–1016

Liu N, Zhong NQ, Wang GL et al (2007) Cloning and functional characterization of PpDBF1 gene encoding a DRE-binding transcription factor from *Physcomitrella patens*. Planta 226:827–838

Lozano R, Giménez E, Cara B et al (2009) Developmental genetics of tomato, a model species in plant biology. Int J Dev Biol (in press)

Lu SY, Jing YX, Shen SH et al (2005) Antiporter gene from *Hordum brevisubulatum* (Trin.) Link and its overexpression in transgenic tobacco. J Integr Plant Biol 47:343–349

Lukens LN, Doebley J (1999) Epistatic and environmental interactions for quantitative trait loci involved in maize evolution. Genet Res 74:291–302

Lv S, Zhang KW, Gao Q et al (2008) Overexpression of an H^+-PPase gene from *Thellungiella halophila* in cotton enhances salt tolerance and improves growth and photosynthetic performance. Plant Cell Physiol 49:1150–1164

Maathuis FJM (2006) The role of monovalent cation transporters in plant responses to salinity. J Exp Bot 57:1137–1147

Maheswaran M, Huang N, Sreerangasamy SR et al (2000) Mapping QTL associated with days to flowering and photoperiod sensitivity in rice (*Oryza sativa* L.). Mol Breed 6:145–155

Manneh B, Stam P, Struik PC et al (2007) QTL-based analysis of genotype-by-environment interaction for grain yield of rice in stress and non-stress environments. Euphytica 156:213–226

Mansour MMF, Salama KHA (2004) Cellular basis of salinity tolerance in plants. Environ Exp Bot 52:113–122

Martínez-Atienza J, Jiang X, Garciadeblas B et al (2007) Conservation of the salt overly sensitive pathway in rice. Plant Physiol 143:1001–1012

Martinez-Rodriguez MM, Estañ MT, Moyano E et al (2008) The effectiveness of grafting to improve salt tolerance in tomato when an 'excluder' genotype is used as scion. Environ Exp Bot 63:392–401

Matsushita N, Matoh T (1991) Characterization of Na^+ exclusion mechanisms of salt-tolerant reed plants in comparison with salt-sensitive rice plants. Physiol Plant 83:170–176

Mikiko LK, Levesley A, Koebner RMD et al (2001) Quantitative trait loci for component physiological traits determining salt tolerance in rice. Plant Physiol 125:406–422

Mittova V, Tal M, Volokita M et al (2002) Salt stress induces up-regulation of an efficient chloroplast antioxidant system in the salt-tolerant wild tomato species *Lycopersicon pennellii* but not in the cultivated species. Physiol Plant 115:393–400

Moller IS, Tester M (2007) Salinity tolerance of Arabidopsis: a good model for cereals?. Trends Plant Sci 12:534–540

Monforte AJ, Asins MJ, Carbonell E (1997) Salt tolerance in Lycopersicon species. VI. Genotype-by-salinity interaction in quntitaive trait loci detection: constitutive and response QTLs. Theor Appl Genet 95:706–713

Moradi F, Ismail AM, Gregorio GB et al (2003) Salinity tolerance of rice during reproductive development and association with tolerance at the seedling stage. Indian J Plant Physiol 8:105–116

Morgan JM (1992) Osmotic components and properties associated with genotypic differences in osmorregulation in wheat. Aust J Plant Physiol 19:67–7676

Moya JL, Gomez-Cadenas A, Primo-Millo E et al (2003) Chloride absorption in salt-sensitive Carrizo citrange and salt-tolerant Cleopatra mandarin citrus rootstocks is linked to water use. J Exp Bot 54:825–833

Munns R, Hare RA, James RA, Rebetzke GJ (2000) Genetic variation for improving the salt tolerance of durum wheat. Aust J Agric Res 51:69–74.

Munns R (2002) Comparative physiology of salt and water stress. Plant Cell Environ 25:239–250

Munns R (2005) Genes and salt tolerance: bringing them together. New Phytol 167:645–663

Munns R, Termaat A (1986) Whole-plant responses to salinity. Aust J Plant Physiol 13:143–160

Munns R, Tester M (2008) Mechanisms of salinity tolerance. Annu Rev Plant Biol 59:651–681

Muñoz-Mayor A, Pineda B, Garcia-Abellán JO et al (2008) The HAL1 function on Na^+ homeostasis is maintained over time in salt-treated transgenic tomato plants, but the high reduction of Na^+ in leaf is not associated with salt tolerance. Physiol Plant 133:288–297

Murthy M, Tester M (1996) Compatible solutes and salt tolerance: misuse of transgenic tobacco. Trends Plant Sci 1:294–295

Nagata T, Iizumi S, Satoh K et al (2008) Comparative molecular biological analysis of membrane transport genes in organisms. Plant Mol Biol 66:565–585

Nakashima K, Tran LSP, Nguyen DV et al (2007) Functional analysis of a NAC-type transcription factor OsNAC6 involved in abiotic and biotic stress-responsive gene expression in rice. Plant J 51:617–630

Neumann P (1997) Salinity resistance and plant growth revisited. Plant Cell Environ 20:1193–1198

Niknam SR, McComb J (2000) Salt tolerance screening of selected Australian woody species - a review. Forest Ecol Manage 139:1–19

Ogawa D, Yamaguchi K, Nishiuchi T (2007) High-level overexpression of the Arabidopsis HsfA2 gene confers not only increased thermotolerance but also salt/osmotic stress tolerance and enhanced callus growth. J Exp Bot 58:3373–3383

Ohta M, Hayashi Y, Nakashima A et al (2002) Introduction of a Na^+/H^+ antiporter gene from *Atriplex gmelini* confers salt tolerance to rice. FEBS Lett 532:279–282

Owttrim GW (2006) RNA helicase and abiotic stress. Nucleic Acids Res 34:3220–3230

Pardo JM, Cubero B, Leidi EO et al (2006) Alkali cation exchangers: roles in cellular homeostasis and stress tolerance. J Exp Bot 57:1181–1199

Park EJ, Jeknic Z, Pino MT et al (2007) Glycinebetaine accumulation is more effective in chloroplasts than in the cytosol for protecting transgenic tomato plants against abiotic stress. Plant Cell Environ 30:994–1005

Penna S (2003) Building stress tolerance through over-producing trehalose in transgenic plants. Trends Plant Sci 8:355–357

Perez-Alfocea F, Estañ MT, Santa Cruz A et al (1993) Effects of salinity on nitrate, total nitrogen, soluble protein and free amino acid levels in tomato plants. J Hort Sci 68:1021–1027

Perez-Alfocea F, Santa Cruz A, Guerrier G et al (1994) NaCl stress-induced organic solute changes on leaves and calli of *Lycopersicon esculentum*, *L. pennellii* and their interspecific hybrid. J Plant Physiol 143:106–111

Pineda B (2005) Análisis functional de diversos genes relacionados con la tolerancia a la salinidad y el estrés hídrico en plantas transgénicas de tomate (*Lycopersicon esculentum* Mill). Ph.D. thesis. Universidad Politécnica de Valencia, Valencia, Spain

Platten JD, Cotsaftis O, Berthomieu P et al (2006) Nomenclature for HKT transporters, key determinants of plant salinity tolerance. Trends Plant Sci 11:372–374

Qiao WH, Zhao XY, Li W et al (2007) Overexpression of AeNHX1, a root-specific vacuolar Na^+/H^+ antiporter from *Agropyron elongatum*, confers salt tolerance to *Arabidopsis* and *Festuca* plants. Plant Cell Rep 26:1663–1672

Quintero FJ, Ohta M, Shi H et al (2002) Reconstitution in yeast of the Arabidopsis SOS signaling pathway for Na+ homeostasis. Proc Natl Acad Sci USA 99:9061–9066

Rajagopal D, Agarwal P, Tyagi W et al (2007) *Pennisetum glaucum* Na^+/H^+ antiporter confers high level of salinity tolerance in transgenic *Brassica juncea*. Mol Breed 19:137–151

Ravens JA (1985) Regulation of pH and generation of osmolarity in vascular plants - a cost-benefit analysis in relation to efficiency of use of energy, nitrogen and water. New Phytol 101:25–77

Rhodes D, Hanson AD (1993) Quaternary ammonium and tertiary sulfonium compounds in higher-plants. Annu Rev Plant Physiol Plant Mol Biol 44:357–384

Rios G, Ferrando A, Serrano R (1997) Mechanisms of salt tolerance conferred by overexpression of the *HAL1* gene in *Saccharomyces cerevisiae*. Yeast 13:515–528

Rodriguez-Navarro A (2000) Potassium transport in fungi and plants. Biochim Biophys Acta 1469:1–30

Rohila JS, Jainb RK, Wu R (2002) Genetic improvement of Basmati rice for salt and drought tolerance by regulated expression of a barley Hva1 cDNA. Plant Sci 163:525–532

Romero C, Belles JM, Vaya JL et al (1997) Expression of the yeast trehalose-6-phosphate synthase gene in transgenic tobacco plants: pleiotropic phenotypes include drought tolerance. Planta 201:293–297

Rumbaugh MD, Pendery BM, James DW (1993) Variation in the salinity tolerance of strawberry clover (*Trifolium fragiferum* L.). Plant Soil 153:265–271

Rus A, Lee BH, Muñoz-Mayor A et al (2004) AtHKT1 facilitates Na^+ homeostasis and K^+ nutrition in planta. Plant Physiol 136:2500–2511

Rus AM, Estañ MT, Gisbert C et al (2001) Expressing the yeast HAL1 gene in tomato increases fruit yield and enhances K^+/Na^+ selectivity under salt stress. Plant Cell Environ 24:875–880

Sairam RK, Tyagi A (2004) Physiology and molecular biology of salinity stress tolerance in plants. Curr Sci 86:407–421

Sakamoto A, Alia, Murata N (1998) Metabolic engineering of rice leading to biosynthesis of glycinebetaine and tolerance to salt and cold. Plant Mol Biol 38:1011–1019

Sanan-Mishra N, Phan XH, Sopory SK et al (2005) Pea DNA helicase 45 overexpression in tobacco confers high salinity tolerance without affecting yield. Proc Natl Acad Sci USA 102:509–514

Sanchez DH, Siahpoosh MR, Roessner U et al (2008) Plant metabolomics reveals conserved and divergent metabolic responses to salinity. Physiol Plant 132:209–219

Santa-Cruz A, Acosta M, Rus A et al (1999) Short-term salt tolerance mechanisms in differentially salt tolerant tomato species. Plant Physiol Biochem 37:65–71

Sax K (1923) The association of size differences with seed coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8:552–560

Seong ES, Cho HS, Choi D et al (2007) Tomato plants overexpressing CaKR1 enhanced tolerance to salt and oxidative stress. Biochem Biophys Res Commun 363:983–988

Serrano R, Culiañez-Macia FA, Moreno V (1999) Genetic engineering of salt and drought tolerance with yeast regulatory genes. Sci Hort 78:261–269

Shannon MC, Gronwald JW, Tal M (1987) Effects of salinity on growth and accumulation of inorganic ions in cultivated and wild tomato species. J Am Soc Hort Sci 112:416–423

Shi HZ, Zhu JK (2002) Regulation of expression of the vacuolar Na^+/H^+ antiporter gene *AtNHX1* by salt stress and abscisic acid. Plant Mol Biol 50:543–550

Shi HZ, Lee BH, Wu SJ et al (2003) Overexpression of a plasma membrane Na^+/H^+ antiporter gene improves salt tolerance in Arabidopsis thaliana. Nat Biotechnol 21:81–85

Shi HZ, Quintero FJ, Pardo JM et al (2002) The putative plasma membrane Na1/H1 antiporter SOS1 controls long-distance Na1 transport in plants. Plant Cell 14:465–477

Shirasawa K, Takabe T, Takabe T et al (2006) Accumulation of glycinebetaine in rice plants that overexpress choline monooxygenase from spinach and evaluation of their tolerance to abiotic stress. Ann Bot 98:565–571

Sickler CM, Edwards GE, Kiirats O et al (2007) Response of mannitol-producing *Arabidopsis thaliana* to abiotic stress. Funct Plant Biol 34:382–391

Sohn KH, Lee SC, Jung HW et al (2006) Expression and functional roles of the pepper pathogen-induced transcription factor RAV1 in bacterial disease resistance, and drought and salt stress tolerance. Plant Mol Biol 61:897–915

Springer PS (2000) Gene traps: tools for plant development and genomics. Plant Cell 12:1007–1020

Su J, Wu R (2004) Stress-inducible synthesis of proline in transgenic rice confers faster growth under stress conditions than that with constitutive synthesis. Plant Sci 166:941–948

Su J, Hirji R, Zhang L et al (2006) Evaluation of the stress-inducible production of choline oxidase in transgenic rice as a strategy for producing the stress-protectant glycine betaine. J Exp Bot 57:1129–1135

Sun YX, Wang D, Bai YL et al (2006) Studies on the overexpression of the soybean GmNHX1 in *Lotus corniculatus*: the reduced Na+ level is the basis of the increased salt tolerance. Chin Sci Bull 51:1306–1315

Sunkar R, Zhu JK (2004) Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. Plant Cell 16:2001–2019

Takehisa H, Shimodate T, Fukuta Y et al (2004) Identification of quantitative trait loci for plant growth of rice in paddy field flooded with salt water. Field Crops Res 89:85–95

Tanksley SD (2004) The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. Plant Cell 16:181–189

Tester M, Davenport R (2003) Na^+ tolerance and Na^+ transport in higher plants. Ann Bot 91:503–527

Tian L, Huang C, Yu R et al (2006) Overexpression AtNHX1 confers salt-tolerance of transgenic tall fescue. African J Biotechnol 5:1041–1044

Tonsor SJ, Alonso-Blanco C, Koornneef M (2005) Gene function beyond the single trait: natural variation, gene effects, and evolutionary ecology in *Arabidopsis thaliana*. Plant Cell Environ 28:2–20

Tseng MJ, Liu CW, Yiu JC (2007) Enhanced tolerance to sulfur dioxide and salt stress of transgenic Chinese cabbage plants expressing both superoxide dismutase and catalase in chloroplasts. Plant Physiol Biochem 45:822–833

Tuteja N (2007) Mechanisms of high salinity tolerance in plants. Meth Enzymol 428:419–438

Verma D, Singla-Pareek SL, Rajagopal D et al (2007). Functional validation of a novel isoform of Na^+/H^+ antiporter from *Pennisetum glaucum* for enhancing salinity tolerance in rice. J Biosci 32:621–628

Vij S, Tyagi AK (2007) Emerging trends in the functional genomics of the abiotic stress response in crop plants. Plant Biotechnol J 5:361–380

Villalta I, Reina-Sánchez A, Bolarín MC et al (2008) Genetic analysis of Na+ and K+ concentrations in leaf and stem as physiological components of salt tolerance in tomato. Theor Appl Genet 116:869–880

Waditee R, Bhuiyan NH, Hirata E et al (2007) Metabolic engineering for betaine accumulation in microbes and plants. J Biol Chem 282:34185–34193

Wang J, Zuo K, Wu W et al (2004) Expression of a novel antiporter gene from *Brassica napus* resulted in enhanced salt tolerance in transgenic tobacco plants. Biol Plant 48:509–515

Wang W, Vinocur B, Altman A (2003) Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. Planta 218:1–14

Wang WQ, Li Y, Zhang YY et al (2007) Comparative expression analysis of three genes from the *Arabidopsis* vacuolar Na^+/H^+ antiporter (AtNHX) family in relation to abiotic stresses. Chin Sci Bull 52:1754–1763

Weckwertha W (2008) Integration of metabolomics and proteomics in molecular plant physiology – coping with the complexity by data-dimensionality reduction. Physiol Plant 132:176–189

Wu CA, Yang GD, Meng QW et al (2004) The cotton *GhNHX1* gene encoding a novel putative tonoplast Na^+/H^+ antiporter plays an important role in salt stress. Plant Cell Physiol 45:600–607

Wu YY, Chen QJ, Chen M et al (2005) Salt-tolerant transgenic perennial ryegrass (*Lolium perenne* L.) obtained by *Agrobacterium tumefaciens*-mediated transformation of the vacuolar Na^+/H^+ antiporter gene. Plant Sci 169:65–73

Xue ZY, Zhi DY, Xue GP et al (2004) Enhanced salt tolerance of transgenic wheat (*Tritivum aestivum* L.) expressing a vacuolar Na^+/H^+ antiporter gene with improved grain yields in saline soils in the field and a reduced level of leaf Na^+. Plant Sci 167:849–859

Yamaguchi T, Blumwald E (2005) Developing salt-tolerant crop plants: challenges and opportunities. Trends Plant Sci 10:615–620

Yang AF, Duan XG, Gu XF et al (2005) Efficient transformation of beet (*Beta vulgaris*) and production of plants with improved salt-tolerance. Plant Cell Tissue Organ Cult 83:259–270

Yang L, Tang R, Zhu J et al (2008) Enhancement of stress tolerance in transgenic tobacco plants constitutively expressing AtIpk2b, an inositol polyphosphate 6-/3-kinase from *Arabidopsis thaliana*. Plant Mol Biol 66:329–343

Yeo AR (1983) Salinity resistance: physiologies and prices. Physiol Plant 58:214–222

Yeo AR (1992) Variation and inheritance of sodium transport in rice. Plant Soil 146:109–116
Yeo AR (2007) Salinity. In: Yeo AR, Flowers TJ (eds) Plant solute transport. Blckwell, Oxford
Yeo AR, Flowers TJ (1986) Ion transport in *Suaeda maritima*: its relation to growth, and implications for the pathway of radial transport of ions across the root. J Exp Bot 37:143–159
Yeo AR, Yeo, ME, Flowers SA et al (1990) Screening of rice (Oryza sativa L.) genotypes for physiological characters contributing to salinity resistance, and their relationship to overall performance. Theor Appl Genet 79:377–384
Yin XY, Yang AF, Zhang KW et al (2004) Production and analysis of transgenic maize with improved salt tolerance by the introduction of AtNHX1 gene. Acta Bot Sin 46:854–861
Yokoi S, Quintero FJ, Cubero B et al (2002) Differential expression and function of *Arabidopsis thaliana NHX* Na^+/H^+ antiporters in the salt stress response. Plant J 30:529–539
Zhang HX, Blumwald E (2001) Transgenic salt-tolerant tomato plants accumulate salt in foliage but not in fruit. Nat Biotechnol 19:765–768
Zhang HX, Hodson JN, Williams JP et al (2001) Engineering salt-tolerant *Brassica* plants: characterization of yield and seed oil quality in transgenic plants with increased vacuolar sodium accumulation. Proc Natl Acad Sci USA 98:12832–12836
Zhang J, Nguyen HT, Blum A (1999) Genetic analysis of osmotic adjustment in crop plants. J Exp Bot 50:291–302
Zhang Y, Wang Z, Xu J (2007) Molecular mechanism of dehydrin in response to environmental stress in plant. Prog Nat Sci 17:237–246
Zhao F, Zhang H (2006) Salt and paraquat stress tolerance results from co-expression of the Suaeda salsa glutathione S-transferase and catalase in transgenic rice. Plant Cell Tissue Organ Cult 86:349–358
Zhao F, Guo S, Zhang H et al (2006a) Expression of yeast SOD2 in transgenic rice results in increased salt tolerance. Plant Sci 170:216–224
Zhao F, Wang Z, Zhang Q et al (2006b) Analysis of the physiological mechanism of salt-tolerant transgenic rice carrying a vacuolar Na^+/H^+ antiporter gene from *Suaeda salsa*. J Plant Res 119:95–104
Zhao F, Zhang XJ, Li PH et al (2006c) Co-expression of the *Suaeda salsa* SsNHX1 and *Arabidopsis* AVP1 confer greater salt tolerance to transgenic rice than the single SsNHX1. Mol Breed 17:341–353
Zhao J, Zhi D, Xue Z et al (2007) Enhanced salt tolerance of transgenic progeny of tall fescue (*Festuca arundinacea*) expressing a vacuolar Na^+/H^+ antiporter gene from Arabidopsis. J Plant Physiol 164:1377–1383
Zhou SF, Chen XY, Xue XN et al (2007) Physiological and growth responses of tomato progenies harboring the betaine alhyde dehydrogenase gene to salt stress. J Integr Plant Biol 49:628–637
Zhu JK (2001) Plant salt tolerance. Trends Plant Sci 6:66–71

Chapter 17
DNA Microarray as Part of a Genomic-Assisted Breeding Approach

Eva Vincze and Steve Bowra

Abstract In the struggle to achieve global food security, crop breeding retains an important role in crop production. A current trend is the diversification of the aims of crop production, to include an increased awareness of aspects and consequences of food quality. The added emphasis on food and feed quality made crop breeding more challenging and required a combination of new tools. We illustrate these concepts by taking examples from barley, one of the most ancient of domesticated grains with a diverse profile of utilisation (feed, brewing, new nutritional uses). Genomic-assisted breeding (GAB) is the 'umbrella' term used to describe a suite of tools now being applied to plant breeding. In the context of genomic-assisted breeding, we will briefly discuss in the second section of this chapter the molecular genetic-based tools underpinning GAB (understanding gene expression, candidate gene selection, allelic complement, quantitative trait loci [QTLs] and fine mapping). The subject of the third section is the use of DNA microarray as a potentially important tool in crop improvement. This section includes a discussion about what can we expect using the DNA microarray technology and what could be major considerations when the technique is applied. We consider the use of cDNA vs. oligonucleotide microarrays, target purification, labelling, hybridisation, image acquisition, minimising random errors, experimental design, biological and technical variability, quality control, normalisation, statistical and practical significances, fold changes, validation and possible additional regulatory mechanisms in gene expression. The subject of the fourth section is the applications of DNA microarrays to study of global gene expression during grain filling in monocot crops, especially barley. We compare large arrays vs. tissue/pathway specific approaches using an example of focused microarray and how it follows predicted changes during grain development. We describe of an extension of the study to field grown material and conclude that such an approach is

E. Vincze (✉)
Department of Genetics and Biotechnology, Faculty of Agricultural Sciences, Aarhus University, Research Centre Flakkebjerg, DK-4200, Slagelse, Denmark
e-mail: Eva.Vincze@agrsci.dk

S. Bowra
Verzyme (UK) Ltd., Plas Gogerddan, Aberystwyth, Wales SY23 3EB, UK

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_17,

able to interpret differences in the gene expression profiles of barley storage protein homologues. Therefore, microarray analysis could provide the knowledge required designing an improved amino acid profile with the possibility of breeding selectively for specific alleles/homologues to confer enhanced amino acid profile of the barley storage proteins and we outline the potential of microarray as a tool to support genomic-assisted breeding approach to improve the nutritional quality of barley.

17.1 Introduction

Barley (*Hordeum vulgare* L.) is today the world's fourth major cereal crop (see Bethke et al. 2000). Archaeological remains of barley grains have been found at various sites in the Fertile Crescent (for example Abu Hureyra and Jericho) that indicate that the crop was domesticated around 8000 BC (see Zohary and Hopf 2000). Recent evidence indicates that a second, independent domestication event may have occurred in Central Asia at the eastern edge of the Iranian Plateau (Morrell and Clegg 2007). It is suggested that the Fertile Crescent domestication contributes the majority of the diversity in European and American cultivars, whereas the second domestication contributed most of the diversity found in barley from Central Asia to the Far East. (Morrell and Clegg 2007; Saisho and Purugganan 2007).

The progenitor of cultivated barley (*Hordeum spontaneum* C. Koch.) is still present in its primary habitats in the Fertile Crescent, from Israel and Jordan to south Turkey, Iraqi Kurdistan, and south-western Iran (Harlan and Zohary 1966; Brown et al. 1978). In the same area, *H. spontaneum* also occupies an array of secondary habitats, such as open Mediterranean marquis, abandoned fields, and roadsides. Similar marginal habitats have been more recently colonized by *H. spontaneum* in the Aegean region, South-eastern Iran, and central Asia, including Afghanistan and the Himalayan region (see Zohary and Hopf 2000).

From the wild ancestors of the Fertile Crescent, barley has become arguably the most widely adapted cereal grain species, because it is found at high altitudes and farther into deserts than any other cereal crop. Barley varieties show a large variation in morphological, ecological and agricultural traits, including two row and six row spikes, growth habit, maturity, plant height, grain size, disease resistance etc. As an example, 133 barley cultivars are approved in Germany, ranging from protein-rich winter barley (which is used as fodder) to starch-rich spring barley used for malting.

17.1.1 Current Uses: Feed, Brewing and the Emerging Importance in Food

Today, as stated, barley is grown in many regions throughout the world, in part due to its ability to adapt to less favourable growing conditions. In 2005, 154 million tons of barley was harvested worldwide, with more than half produced in the European Union. Of the total annual barley production, approximately two thirds is

used for feed, one third for malting and a small (2%) but increasing percentage used as food and in food processing.

Barley is primarily incorporated into monogastric feed formulations to provide a source of protein, energy but also fibre, which fosters general well-being through prebiotic activity by selectively stimulating the growth and/or activity of one or a limited number of bacteria in the colon and thus improve host health.

The protein content and quality is one of the major determinants of the nutritional value of barley grains and grain-derived products. However, present day cultivars are characterised by low protein content and deficiencies in essential amino acids (AA) (Table 17.1). The major storage proteins in barley also have significantly high levels of two non-essential AAs, proline and glutamine, which are readily de-aminated during digestion. The excess nitrogen is excreted, resulting in environmental nitrogen load. The impact of the poor protein quality of barley, in particular the amino acid imbalance, is currently addressed through the addition of amino acid supplements, produced via fermentation, to diet formulations. The net result of the 'industrial solution' to poor protein quality is that little research effort has been directed specifically toward improving amino acid and protein composition of barley or other cereals. Therefore, when breeding barley under modern high input agriculture conditions, the most important selection criteria have been yield and high nitrogen content. Although nitrogen utilisation efficiency by livestock has been greatly improved through optimised diet formulation incorporating soybean

Table 17.1 AA concentrations of barley, the high lysine mutant of barley and soybean grains and the ideal AA concentration in pig diet (g/16 g nitrogen.) (Modified from Lasztity 1996; Boisen et al. 2000.)

Amino acids	Barley	High lysine mutant 1508	Soybean	Ideal AA in the diet for growing pig
Lysine	3.5	5.4	5.7	7.1
Histidine	2.1	2.8	2.5	2.5
Arginine	4.4	7.1	7.1	–
Aspartic acid	6.1	8.3	11.3	–
Threonine	3.5	4.3	3.3	4.5
Serine	4.2	4.8	4.6	–
Glutamic acid	24.6	17.0	14.6	–
Proline	10.9	7.3	0.9	–
Glycine	4.2	5.6	5.1	–
Alanine	4.1	5.1	3.6	–
Cysteine	2.5	2.1	1.3	3.6 (+ Met)
Valine	5.4	5.8	4.4	5.2
Methionine	1.6	2.2	1,2	1.8
Isoleucine	3.8	3.6	4.7	4.0
Leucine	6.9	6.9	6.6	8.0
Tyrosine	2.5	3.7	4.0	–
Phenylalanine	5.1	4.4	5.1	8.0 (+ Tyr)
Tryptophan	1.4	1.6	1.1	1.2

meal and essential AA supplements, barley remains a sub-optimal source of protein due to (1) the large excess of proline and glutamine, which is deaminated by the animals and excreted causing environmental pollution (2) the deficiency in essential amino acids (lysine, threonine, methionine, cysteine, leucine and valine) some of which (leucine and valine) are not commercially available.

To date, despite the fact that over the past 30–40 years research has identified significant naturally occurring variation in nutritional quality traits, most attempts to address the quality of the amino acid composition in barley have focused on genetic and molecular manipulations, which have had limited success. Over the last 5 decades, large-scale mutagenesis programs have been devoted to improving the amino acid composition of barley. A range of primarily recessive mutants has been identified that confer a high lysine phenotype (see Munck 1992). However, the high lysine trait is invariably associated with pleiotropic effects that affect yield, quality and agronomic performance. Despite substantial improvements in the yield of *lys* mutants, they are not commercially viable (see Munck 1992). Additional alleles have been found, in particular in the *lys3* locus where the *lys3a* allele confers the highest lysine phenotype (Doll 1980). At present there is a limited understanding of the nature of the different *lys* mutations with the exception of Mutant 56 that has a large deletion covering the locus for the major group of prolamine storage proteins, the B-hordeins (Kreis et al. 1983). In the *lys3a* mutant, there is a decrease in the amount of B and C hordeins due to reduced transcription but an increase in the high molecular weight D-hordeins (Kreis et al. 1984). There are enhanced transcript and protein levels of chymotrypsin inhibitors and a α-amylase/subtilisin inhibitor protein, all of which have high lysine content.

In addition to a mutation approach, the manipulation of the genome using molecular genetic transformation technology has been applied with the aim of improving the seed lysine content in a wide range of species (Galili et al. 2005). The efforts have addressed (1) circumvention of the feedback regulation of lysine and threonine biosynthesis in the aspartate family pathway by overexpressing bacterial genes encoding feedback-insensitive aspartate kinase and dihydrodipiciolinate synthase; (2) reducing lysine catabolism and (3) the introduction of genes encoding heterologous storage proteins with for example high of levels lysine and methionine including synthetic proteins designed with high lysine and methionine content. However, although significant increase in free lysine content has been achieved, but other drawbacks have been reported such as reduced seed viability.

The application of anti-sense or RNAi technology has illustrated the potential to modify the proportion of storage proteins in a grain. In rice, it has been shown that anti-sense suppression of glutelin synthesis results in an increased accumulation of prolamins that confers improved functional properties with regard to the brewing of sake (Maruta et al. 2002). In oilseed rape (*Brassica napus)*, Kohno-Murase et al. (1995) suppressed the synthesis of nutritionally poor storage proteins, cruciferin, by anti-sense technology and reported a corresponding increase in napin, the other major storage protein, which have a proportionally higher

lysine and methionine content. In maize, RNAi inhibition of the 22 kD α-zeins transcript resulted in an 18.5% increase in lysine content, with limited changes in the abundance of other storage proteins (Segal et al. 2003). Similarly targeting the 19 kD α-zein gene family using an anti-sense construct led to significant increases in lysine, threonine and tryptophan (Huang et al. 2004). To date there has only been one reported modification of the hordein composition in barley (Lange et al. 2007). Using anti-sense technology targeted towards C-hordeins, the prolamine with the highest mole percentage of proline and glutamine, resulted in a reduction of the C-hordein content and relative increase in the B-, D- and γ-hordeins as well as increases in the amounts of albumins/globulins and glutelins. Overall, the induced changes resulted in decreased proline, glutamine and phenylalanine levels and increased amounts of the remaining amino acids (Lange et al. 2007).

In spite that barley is predominately used as feed, a significant proportion of barley plant breeding has focused on malting quality. Sparrow (1970) defined malting as "the commercial exploitation of those processes that lead to germination". Problems in defining and analysing malting quality derived from the complex interactions of a very wide range of characters are reinforced by the estimated 150 quantitative trait loci that interact and influence malting quality. However, there are four major aspects of the malting process that can be considered: hormone control, production of enzymes, the substrates like starch, protein and cell wall within the endosperm, and the transport of soluble materials to the embryo. In short, malting quality is a complex trait and the "ideal" malting barley remains to be described.

Barley is thought to have been first used for food, but evolved primarily into a feed, malting and brewing grain due, in part, to the rise in prominence of wheat and rice. In spite of the limited current use of barley as food (2% of the word production), throughout history, barley has remained a major food source for some cultures, principally in Asia and northern Africa (Newman and Newman 2006).

Today, it is increasingly recognised that the major advantage of incorporating barley into various food products stems from its potential health benefits (Baik and Ullrich 2008). It is widely accepted that barley β-glucans are effective in lowering blood cholesterol level and decrease diabetic risks (Pins and Kaur 2006). In 2005, the American National Barley Foods Council submitted a health claim petition to the Food and Drug Administration of the USA suggesting that the increase in beta-glucan soluble fibre (from de-hulled or hulless whole or pearl barley, barley flakes, barley grits, barley meal, barley flour, barley bran, and beta-glucan enriched barley meal or flour) could prevent coronary heart disease. The Food and Drug Administration of the USA (Food and Drug Administration HHS 2008) ratified the claim. Barley is also a rich source of tocols, including tocopherols and tocotrienols, which are known to reduce serum LDL cholesterol (Panfili et al. 2008). Therefore, in summary barley as a food and nutraceutical continues to attract considerable attention from the food industry along with greater consumer awareness of the role of food quality in supporting wellness and the quality of life.

17.1.2 Breeding for Quality Traits

The impact of global climate change on agricultural production is both direct, as a result of increasing temperatures and more erratic rainfall (Parry et al. 2005), and indirect, resulting from the pressure to produce additional yield and bioenergy, while minimising carbon inputs into crop production (Powlson et al. 2005). There is therefore an urgent need to breed new crop varieties that have both improved yield (Parry et al. 2007) and improved nutrient content (Hirel et al. 2007). This is in fact nothing new and was addressed during the advent of the so-called 'green revolution'. However, what is new is the need to understand the impact of nutritional quality of cereals, in particular barley, and the potential to improve animal and human nutrition and overall wellness. To this end, genomics-based technologies, which hold very significant potential to assist focused barley breeding programmes, need to be efficiently implemented.

17.2 Genetic Analysis: Combining Methods

In the previous section, it was concluded that plant breeding remains essential and breeding for complex traits, such as nutritional quality, are ever more important given the tangible link between food and feed quality and overall health. Many crop traits are quantitative, complex and controlled by multiple, interacting genes. Advances in genomics have made it possible to analyse the architecture of cereal genomes to assign quantitative traits loci to key agronomical traits. The advent of molecular markers combined with the progress in quantitative traits loci mapping has resulted in new breeding approaches such as marker-assisted selection and 'breeding by design' (Ribaut and Hoisington 1998; Peleman and van der Voort 2003).

With an increase in the number of genomes which have been sequenced and the large and ever-growing amount of expressed sequence tag (EST) sequences available, we are entering the post genomics era where the need to identify and apportion gene function becomes the focus. Therefore, moving from the genomic era into a post genomic era ushers in functional genomics and systems biology. Both approaches combine a wide range of technologies designed to collect, integrate and synthesis data derived from genetic, proteomic, metabolomic and physiological analysis to provide a platform to understand gene function in the context of the genome and environment.

17.2.1 Genomic-Assisted Breeding

The emergence of the post genomic era has seen the rapid evolution of high throughput technologies, which are designed to support the study of multiple genes (DNA microarray), the gene products (protein array, proteomics) and overall

impact on metabolic profile (metabolomics). The integration of these technology platforms offers significant potential to analyse genes and gene products in parallel and reveal complex regulatory events. Further insights may be gained from gene expression values from microarray profiling as genetic studies of gene expression have treated transcript abundance itself as a quantitative trait and have mapped it to local-acting or distant-acting expression quantitative trait loci (eQTLs) (Brem et al. 2002; Zhang et al. 2007). The combination of molecular and genetic analysis can unravel the genetic network and elucidate epistasis, which is often found in phenotypes with interactive and interrelated metabolic and ontogenetic pathways (Lee 1995). Overlaying the combination of molecular and genetic studies with an understanding of plant metabolism regulation can provide new tools for breeders to increase crop yields (Morandini and Salamini 2003). Therefore by combining conventional breeding protocols with established and emerging genomics-based approaches (functional genomics, quantitative loci analysis and marker assisted breeding), the concept of genomic-assisted breeding (GAB) has evolved. Genomic-assisted breeding provides the opportunity to enhance the prediction of the phenotype from a genotype for manipulation and design of complex traits supporting enhanced plant breeding (Varshney et al. 2007).

17.2.2 Molecular Genetic-Based Tools Underpinning GAB

Genetics and genetic analysis has long been the basis of traditional plant breeding, which has been and continues to be successful. As mentioned above, DNA marker technology, which emerged during the 1980s, offers great promise for plant breeding. Owing to genetic linkage, DNA markers can be used to detect the presence of allelic variation in the genes underlying these traits. The application of DNA markers has fostered the development of marker assisted (MAS) plant breeding, which can support increased efficiency and precision, although to date results are inconsistent (Hospital 2008).

The first DNA markers were developed from genomic DNA probes which were completely random and not characterised. The increased availability of sequence data for genes through genome and EST-sequencing projects has lead to the next generation of DNA markers, called 'functional' markers or expression based markers (Andersen and Lubberstedt 2003). Functional markers have significant advantages over random markers because they can be linked to the desired trait allele and therefore the associated phenotype. Functional markers can be derived from the gene responsible for the trait of interest and target the functional polymorphism in the gene, thus allowing its selection in different genetic backgrounds without re-validating the marker-QTL allele relationship (Varshney et al. 2005a). In addition to being useful for identifying the 'perfect' or 'ideal' markers in marker-assisted selection, functional markers are an important resource for estimating functional variation in natural or breeding populations and for studying genome evolution, through comparative mapping (Varshney et al. 2005b).

In summary, genomic-assisted breeding encompasses and extensive range of tools, techniques and approaches, which in concept, can produce an enhanced picture of the genome-phenotype relationship (Fig. 17.1). It is beyond the scope of this chapter to go into more detail offered in other chapters within this volume. Therefore the following is an overview of the tools and technologies related to the creation and utilisation of functional markers and how these will enable GAB.

17.2.2.1 Understanding Gene Expression

Functional markers are synonymous with expression based markers and are distinguished from genomic DNA markers on the basis that they are derived from expressed sequence information. More specifically functional markers can be

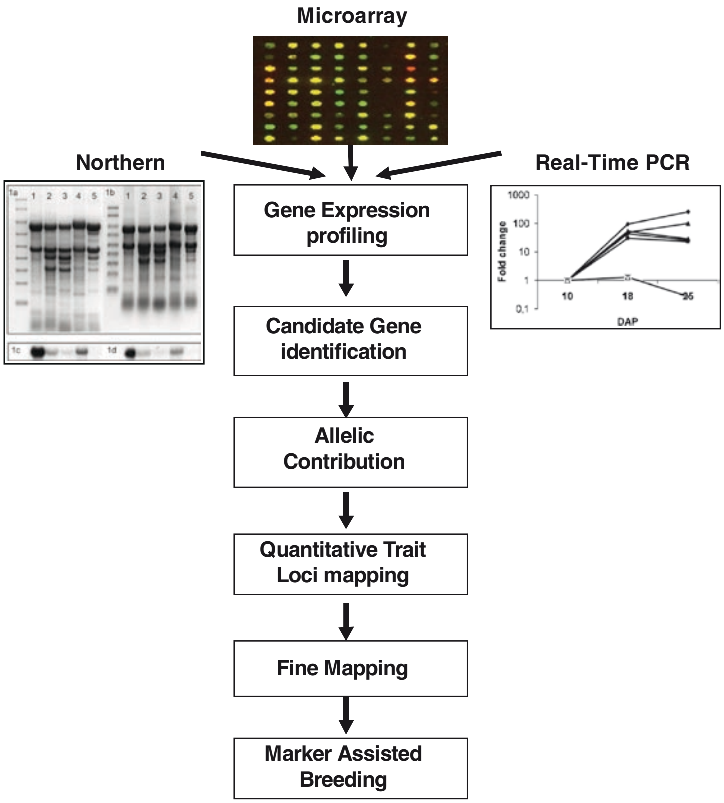


Fig. 17.1 The components of genomic-assisted breeding, the figure is presented as series of technical steps which would be under taken in practise to achieve the functional markers that would then be applied in a marker assisted breeding programme. The origin of pictures: the left is a Northern Vincze and Bowra (2005), the middle on the top is a microarray results from Hansen (HM personal communication 2007), the right is a real time PCR from Hansen et al. (2009)

associated with alleles which confer phenotypic variation within both natural land races/germplasm and breeding lines. Information derived from the functional marker association can contribute to the understanding of important complex agronomical traits when placed in the context of the species ploidy, level of gene duplication (gene families), epigenesis and gene expression. Stripping away the discussion of epigenetics, ploidy etc. reveals that transcriptional activity of a gene is the underlying genetic component. As a consequence of the importance of transcription information, there is a need for ways of reliably determining, quantifying, presenting and integrating such information which in turn provides the building blocks for functional marker development. Therefore understanding the temporal, spatial and developmental expression profile of a gene or genes is an important first step towards creating markers which are linked to the underlying metabolism responsible for a phenotype/trait.

Determining the steady state level of mRNA provides a snapshot of the accumulation level of a gene transcript according to species, tissue type and developmental stage. The recent advent of DNA microarray has opened the opportunity to assess the steady state level of many genes in parallel and therefore improves on Northerns and RT-PCR in terms of numbers that can be handled easily Using microarray as a basis for gene expression analysis provides the opportunity to integrate expression data from genes coding for proteins associated with specific metabolic pathways and draw correlative interactions between genes which could confer synergy within the transcriptome (first step in Fig. 17.1; the selection of the method to study gene expression).

17.2.2.2 Candidate Gene Selection

Expression information derived from microarray is a powerful step towards generating a more complete picture of gene expression and therefore by correlation, the gene product. The utility of the derived information is made more relevant if a targeted pathway specific analysis is applied (see Sects. 17.3 and 17.4). DNA microarray can suggest target genes which, through the action of the gene product, can confer significant influence on the plant phenotype under specified environmental conditions, where this phenotype is observed. The gene or genes identified to confer a phenotype, are called candidate genes as shown in step 2 in Fig. 17.1, and provide the basis for establishing targeted functional markers which can be integrated into mapping and marker assisted programmes.

17.2.2.3 Allelic Complement

Drawing on studies from the field of human diseases suggests that genetic variation affecting gene expression has significant effects on phenotypic variability. Variation *in Cis*-acting factors that confer regulation leads to preferential expression of an allelic transcript, and can be detected *in vivo* by comparisons of the relative abundance of these allelic transcripts using intragenic polymorphisms (Pastinen and

Hudson 2004). Such surveys of allelic expression have established that unequal representation of marker alleles in transcripts from heterozygous individuals is common and occurs in approximately 5–20% of heterozygous individuals, and in 15–50% of their genes (Pastinen et al. 2004). Allelic expression analysis may also detect epigenetic regulation, genomic imprinting and random mono-allelic expression (Pastinen and Hudson 2004). In plants, hybridization is a common process and it plays a vital role in plant breeding. Hybridization can generate phenotypic novelty, including a broad array of new and sometimes transgressive phenotypes (Rieseberg et al. 2003). Intra-specific hybridization between two cultivars, ecotypes, or accessions can also result in up- or down regulation of gene expression, for examples, in diploid and triploid maize hybrids (Stupar et al. 2008 and references therein) in diploid wheat (Wu et al. 2003; Wang et al. 2006) and rice hybrids (Bao et al. 2005). Differences in allelic expression of non-imprinted autosomal genes occur in intra-specific maize hybrids and show unequal expression of the parental alleles (Guo et al. 2006; Stupar et al. 2007). Therefore, the knowledge of allelic variability, including expression patterns, is a vital element in genomic-assisted breeding and forms step 3 in Fig. 17.1.

17.2.2.4 Quantitative Trait Loci (QTLs) and Fine Mapping

Many important agronomical traits are under the control of multiple genes working synergistically across the genome. Complex traits are referred to as quantitative traits. The association between a complex phenotypic trait and genetic markers on the chromosomes can be detected through statistical analysis, leading to the identification of quantitative trait loci (QTL), chromosome regions that appear to be associated with the given phenotype (step 4 in Fig. 17.1). Quantitative trait loci usually cover a large section of the associated chromosomes, potentially containing many genes, most of which will have no role in conferring the phenotype. One mechanism by which a gene might be associated with the trait is through altered transcription. Microarrays have the ability to measure a large percentage of the genes in the genome, and this assessment parallels the genome-wide scan performed by QTL methods (Han et al. 2008). In recent years there has been increasing number of reports that attempt to combine microarray data with QTL data (Wayne and McIntyre 2002; Alba et al. 2004; Han et al. 2008). This has resulted in the development of a web-based software tool for combined visualization and exploration of gene expression data and QTL in animal science: the GeneNetwork website (WebQTL, http://www.genenetwork.org/). This allows the assessment of the relationship between gene expressions and QTL in recombinant inbred mice (Wang et al. 2003). We suggest that a similar approach offers future potential for plant sciences, however, it is important to stress that comparing QTL and microarray data is not straightforward. The high level of experimental errors and limitations in microarray data analysis, discussed in Sect. 17.3, can introduce mistakes in the identification of relevant genes. In addition, QTL studies include the entire genome, including noncoding regions, while microarray studies seldom do this.

While QTLs are defined through the application of markers, these are non-specific markers and not functional markers. Fine-mapping involves the identification of markers that are very tightly linked to a targeted gene (step 5 in Fig. 17.1). In case when the genes supporting the quantitative trait have been identified, for example using a microarray, these can become functional markers, and in turn can be used to assist the development of genetic maps with a high density of molecular markers. The net result is the development of genetic and physical maps to understand the genome architecture and assist the implementation of molecular-assisted breeding strategies (step 6 in Fig. 17.1).

17.3 Messenger RNA Expression Profiling Using DNA Microarray Technology

DNA microarray has become a standard method for the analysis of the transcriptome, suitable for all biological species. However, drawing conclusions from the large amount of data obtained has been and remains a challenge to the extent that the original excitement surrounding DNA microarray has been replaced by an emerging view the expectation have exceed reality (Walker and Hughes 2008). There have been difficulties with accurate reporting of the results, with experimental reproducibility and with identifying and interpreting the biologically relevant information. In spite of this microarray has significant untapped potential. In this context the following section will focus on the technical and biological issues which currently constrain the successful exploitation of the technology and raise awareness of the value of data and the impact on data interpretation.

17.3.1 The Principle of mRNA Expression Profiling: What Can Be Expected?

Over the last 10 years, a considerable amount of plant genomic data has been generated, including complete genome sequences of *Arabidopsis thaliana*, *Medicago truncatula*, *Orysa sativa*, *Populus trichocarpa* (NCBI Reference Sequence: ftp://ftp.ncbi.nih.gov/genomes/); and many more are in progress. The publicly available data sets have contributed to the rapid development of the DNA microarray technology, which enables a simultaneous analysis of a large numbers (up to 2.1 million) of mRNAs expressed in a single biological sample. The transcriptional profiles generated contain potential information concerning the total cellular of the cell, the principal determinant of protein levels and therefore of cellular phenotype (Lockhart and Winzeler 2000). Interpreting the large volume of data generated by microarray technology poses a number of unusual new challenges; one of them is to develop an explanation of the behaviour of a single gene that is consistent and well supported by complementary information provided by thousands of other genes (Breitling 2006). Therefore it is very important to stress that realising the

true potential of microarray requires appropriate awareness of the sources technical and biologically variation and how to when extracting information from the data generated.

17.3.2 The Development of Array Technology

Gene expression analysis first began with the development of a hybridisation-based technique referred to as ‘Northern blot’ (Alwine et al. 1977). Over the following decades a variety of techniques have emerged based around hybridisation, PCR and or a combination of both, for example quantitative real-time RT-PCR (Nolan et al. 2006), differential display (Liang and Pardee 1992), and suppressive subtractive hybridisation (Diatchenko et al. 1996). However, while producing significant data, these techniques have limited potential for large scale gene analysis. To meet this challenge a number of other technique was developed: serial analysis of gene expression (SAGE) (Velculescu et al. 1995), massive parallel signature sequencing (MPSS) (Brenner et al. 2000), cDNA microarray (Schena et al. 1995), and oligo microarray chip technologies (Lockhart et al. 1996) have emerged as the techniques of choice when addressing global gene expression studies. Therefore in this context our discussion will focus on the use and application of both cDNA and oligo microarray chip technologies.

17.3.2.1 cDNA vs Oligonucleotide Microarrays

DNA microarrays can be fabricated using either complementary DNA (cDNA) clones, derived from cDNA libraries, which have been constructed from mRNA of the target organism or from short synthetic oligonucleotides designed from sequence information derived from genomic or EST sequencing programmes. The cDNA or the oligonucleotides that form the targets for hybridisation are called probes. The choice of DNA microarray will depend on the number of probes under examination, costs, customization requirements, and the type of scientific question being asked. As a result, of the design criteria microarrays can have as little 100 probes to up to 2.1 million micron-scale probes on one chip.

Fabrication of the microarray starts with the preparation of the probes. These can be (a) cDNA sequences derived from EST clones or small fragments of PCR products or (b) synthetic oligonucleotides, which are short sequences designed to complement known expressed sequences, predicted open reading frames or 3’UTRs. The oligo-nucleotides may be long (60-mer) or short (25-mer) probes depending on the desired purpose; longer probes are more specific to individual target genes, shorter probes may be spotted in higher density across the array and are cheaper to manufacture. Complementary cDNA-based probes can be spotted using a robot onto a surface, normally a glass slide coated with a charged polymer (for example Poly-L-lysine, amino-silane or amino-reactive silanes) to provide adhesion. In the case of oligo-nucleotides probes, it is possible to synthesise them directly onto the support surface. In the robotic spotting procedure, the syringe with a capillary tube, to

which a low but constant pressure is applied, spots a few nanoliters of each probe to a number of glass slides. The DNA is diluted in a spotting buffer containing 50% dimethyl sulfoxide (DMSO) or 1.5 M betaine to ensure optimal spot morphology, signal intensity and DNA distribution. After spotting, DNA is immobilised to the matrix by cross-linking with ultraviolet irradiation (200–400 mJ), when covalent bonds are formed between the amine groups of the matrix and the DNA. After fixation, the slides are blocked using either succinic anhydride or bovine albumin (BSA) in 50% formamide. The remaining free amine groups on the surface are reduced by succinic anhydride mixture to avoid any unspecific hybridisation. The BSA solution is also used for blocking free amine groups.

The resulting "grid" of probes provides the address to receive the "target" complementary cDNA or cRNA derived from experimental samples. The technique supports the fabrication of 'in-house' printed microarrays. The fabrication of in-house microarrays are easily customized through the choice of probes and probe locations on the slide. In summary 'in-house' cDNA base microarrays are not expensive and allow flexibility.

However, as can be seen from the description of probe and array preparation, the procedure involves a significant number of steps, all of which can contribute to overall technical variation that should be controlled during the experimental design and should be carefully considered during the data analysis.

17.3.2.2 Target Purification, Labelling, Hybridisation

The DNA probes arrayed on a surface generate a grid reference and associated address to which the 'test' cDNA or cRNA is hybridised. The purity and integrity of the RNA is one of the most crucial factors influencing dye stability and probe-target hybridisation. Consequently, it is essential that potential contaminants such as genomic DNA, cellular proteins, lipids and polysaccharides all of which can affect the signal intensity by non-specific binding to the probe, are removed.

The target cDNA is labelled with a fluorophors to enable the probe address on the microarray to be identified. Two-colour microarrays or two-channel microarrays are typically used with cDNA prepared from two samples to be compared (e.g. stressed versus non-stressed plants) and they are labelled with two different fluorophores. Fluorescent dyes commonly used for cDNA labelling include cyanine3 (Cy3), which has a fluorescence emission wavelength of 570 nm (emits green light), and guanine5 (Cy5) with a fluorescence emission wavelength of 670 nm (emits red light). The labelling is performed using either direct incorporation of fluorescent nucleotides during reverse transcription or a two-step indirect- or amino allyl labelling incorporation method. The direct incorporation of nucleotides labelled with Cy3 and Cy5 is typically low and influenced by the cDNA sequence, thereby negatively affecting cDNA yield. The indirect method is free of this problem, and it is emerging as the preferred protocol circumventing low incorporation rates and dye biases. The indirect method utilizes an amino allyl modified dUTP (instead of the pre-labelled nucleotide), which is coupled to a reactive ester of Cy3 and Cy5 (Yu et al. 2002).

The two Cy-labelled cDNA samples are mixed and hybridized to a single microarray, which is subsequently scanned in a microarray scanner to visualize fluorescence of the two fluorophores after excitation with a laser beam of a defined wavelength. Relative intensities of each fluorophore may then be used in ratio-based analysis to identify up-regulated and down-regulated genes (Tang et al. 2007).

Oligonucleotide microarrays often contain control probes designed to hybridise with internal RNA controls called spike-ins. Spike-ins are *in vitro* synthesised, polyadenylated transcripts, which specifically anneal only to complementary control probes on microarrays, with minimal self- or cross-hybridization. The degree of hybridization between the spike-ins and the control probes is used to normalize the hybridization measurements for the target probes. Although it is suggested that absolute levels of gene expression may be determined using the two-colour array, the relative differences in expression among different spots within a sample and between samples is the preferred method of data analysis for the two-colour system.

As alternative to the two colours array, single-channel microarrays or one-colour microarray experiments can be performed. The apparent benefit of the single-channel arrays is that it offers estimations of the absolute levels of gene expression. Therefore the comparison of two conditions requires two separate single-dye hybridizations. As only a single dye is used, the data collected represent absolute values of gene expression. These may be compared to other genes within a sample or to reference (‘normalizing’) probes used to calibrate data across the entire array and across multiple arrays. One of the strengths of the single-dye system is that an aberrant sample cannot affect the raw data derived from other samples, because each array chip is exposed to only one sample. In a two-colour system, a single low-quality sample may drastically impinge on overall data precision even if the other sample was of high quality. Another benefit is that data are more easily compared to arrays from different experiments; the absolute values of gene expression may be compared between studies conducted months or years apart. A drawback to the one-colour system is that, when compared to the two-colour system, sufficient replication has to be performed to control for between slide and treatment variation.

17.3.2.3 Image Acquisition

In the preceding section, the practical issues relating to array fabrication and the labelling of the target were addressed; the next step is hybridisation. The procedure of hybridising complementary target to slides arrayed with the probes also requires optimisation to minimise background and ensure high-quality signal intensity. Theoretically, the hybridising conditions (such as target concentration, ionic strength of the buffers and temperature) should and can vary according to the melting temperature of the probes that form the array. However, it is usual practise to use a compromise, which indirectly affects the reliability of comparisons between experiments. Following hybridisation, the slides are washed sequentially with buffers of increased stringency e.g. a three step washing procedure; (1) low stringency buffer

wash (2 x SSC, 0.1% SDS), (2) medium stringency wash (0.1 x SSC, 0.1% SDS), (3) (0.1 x SSC) followed by a brief wash in doubled deionised water. After the final wash, the slides are scanned immediately to avoid bias emerging from the greater susceptibility of Cy5 flourophore to photodegradation compared to Cy3. The image is acquired at 633 and 543 nm to excite Cy5 and Cy3, respectively. The results of the scan are used to calculate the relative expression levels of each gene to identify differentially expressed genes.

17.3.3 Minimising Random Errors

17.3.3.1 Experimental Design

The importance of experimental design can never be overstated and microarray studies are no different. In fact, given that microarray is beset with a significant number of sources for technical variation, appropriate experimental design to control for technical and biological variation is paramount (reviewed in Churchill 2002; Chen et al. 2004). However, while it essential to seek the assistance of advice and support of statisticians; this is best done when the actual biological question has been formulated. Although it may sound obvious to state the need to formulate a biological question, but a brief review of the literature suggests that developing a list of differentially expressed genes has been and remains the prime motive for conducting a significant majority of microarray experiments (White and Salamonsen 2005).

17.3.3.2 Biological and Technical Variability

As stated above microarray experiments have multiple sources of both biological and technical variation, each of which must be considered in the experimental design. The amount and type of biological variation will vary according to the organism, growth conditions and overall objective of the experiment. In the specific case of plants, biological variation between individuals and within individual plants is compounded by growth conditions (environment) and differential rates of development. Therefore it is extremely important to design experiments to ensure adequate replication and sample size such the results obtained truly reflects the population under investigation. Having stressed the need to design an experiment where all sources of biological variation are controlled for and data points replicated, reality is that often there are constraints imposed on the optimal experimental design. Microarray experiments are no exception, and indeed given the associated costs of the procedure it likely that care has to be taken not compromise the experiment due to financial constraints.

Microarray as a technique combines multiple steps, which in turn are made up of many manipulations, as result micro array is ‘riddled’ with sources of technical variation. Technical variation arises from RNA extraction, reverse transcription,

label incorporation and hybridisation steps. Differences in hybridization efficiencies can occur between spots, between different print-tip groups across the array and between slides in the same and different print runs which combine to contribute to measurement errors. Technical variation and measurement error can also interact. For example, the scanning properties of the fluorescent dyes can vary with the spot intensity and spatial position on the slide (Smyth and Speed 2003).

17.3.3.3 Quality Control

It is critical that microarray experiments are carefully controlled, particularly when using dual colour fluorescence microarrays in which the endpoint is a ratio of expression between two or more samples. As in any experiment, treatment controls must be carefully incorporated into the study design. To ensure that there is only one source of experimental variation, consistency must also be applied to tissue collection, processing and RNA extraction, as well as the microarray hybridisations. Even with a single variable, such as a differentiation stimulus, it is possible to end up comparing cells or tissues in completely different physiological states. In this situation, differentially expressed genes will likely be the consequence, rather than the cause, of the differences in phenotype. This problem can be minimised by using carefully controlled inducible systems and examining early rather than later time points (White and Salamonsen 2005).

17.3.3.4 Normalisation

Having acquired data, the next and not insignificant challenge is the transformation of the data, to allow relevant information to be obtained. Often, the first essential step is the normalisation of microarray data to adjust for systematic non-biological effects arising from technical variation and measurement error (see ‘Experimental design’ above). The aim of normalisation is to remove the ‘noise’ from the data, while still maintaining the ability to detect significantly differentially expressed genes. Dual colour fluorescence microarrays require normalisation to account for differences between microarrays, print-tips groups and fluorescent dye channels (reviewed in Smyth and Speed 2003). There is no universally accepted method of microarray data normalisation and a description and comparison of all available methods is beyond the scope of this book chapter.

17.3.3.5 Statistical and Practical Significances

The aim of a microarray experiment is usually to identify differentially expressed genes, with a measure of statistical significance (reviewed Dudoit et al. 2002; Cui and Churchill 2003). Most microarray experiments are designed with only

one categorical factor (e.g. treatment or genotype), therefore the paired t-test is widely applied to analyse them. Experiments with multiple categorical factors (e.g. genotype and time) require methods based on the analysis of variance (ANOVA). Once the data are appropriately normalised, it is common practice to apply univariate testing for each gene and calculate the Student's t-statistics (Dudoit et al. 2002), using the null hypothesis that the mean expression levels of the two samples (e.g. treatment and control) equal. However, while statistical analysis is clearly useful it is important to retain a sense of perspective. This point is encapsulated by Kirk (1996) who defined the difference between statistical significance and practical significance: 'Statistical significance is concerned with whether a research result is due to chance or sampling variability; practical significance is concerned with whether the result is useful in the real world'. Therefore, and by way of caution the utility of the microarray data may not be decided by the significance of the differential expression of a gene but by the understanding of the biological relevance of the change, and indeed to what extent gene expression is directly correlated with protein synthesis and thus phenotype.

17.3.3.6 Fold Changes

In the previous section, the utility of statistical analysis was discussed, while emphasising the need to retain the biological context. That said, detecting genes that show reproducible differences in mRNA abundance between sample classes is probably the most basic and important step of microarray analysis. As a direct consequence there are a large number of methods used in an attempt to establish differential gene expression; all have certain virtues and failings. Despite being widely used, a simple calculation of fold-change and apportioning an arbitrary threshold is a flawed approach. A large body of literature indicates the importance of statistical analysis of microarray data, and points out that the risk of misusing fold change to identify differentially expressed genes is substantial (Lee et al. 2000). Furthermore, the fold change ratio of a microarray experiment is not an absolute measure of expression ratios, because it is the sum of both background and signal intensities. Unavoidable differences in affinity between probes will result in differences in apportioned fold change (Kothapalli et al. 2002). Hence, the fold change is only a relative measure and only to some extend reflect the absolute difference in expression between the conditions compared. Therefore fold change will fail to discriminate between changes that are false and those that are reproducible because setting a fold-change threshold is very subjective. Similarly when the number of replicates is very small, the use of methods based on Student's t-test, including the popular "Permutation t-test method", SAM (Tusher et al. 2001), is questionable and should be replaced by equally simple approaches that are designed for small numbers of replicates, such as Rank Products (Breitling et al. 2004; Breitling and Herzyk 2005). Rank Products have the additional benefit that, by design, the

method can detect rare but interesting expression changes (with statistical significance), that would be missed by many other approaches (Breitling 2006).

17.3.4 Validation

Despite the fact that microarray is rapidly becoming a dominant technology, the research community does not yet accept microarray data without independent validation (reviewed in Rockett and Hellmann 2004). There are a number of reasons for this caution, including the relatively recent development of the technology, the lack of standard operating procedures, the potential for errors in the data and the already mentioned potential biological and technical errors during the performance of a microarray experiment (Kothapalli et al. 2002; White and Salamonsen 2005).

Different techniques can be considered for microarray validation: Northern blot, RNAse protection assay and quantitative real-time RT-PCR (qRT-PCR). The advantage of Northern blot and RNase protection assay is that they provide a semi-quantitative measure as well as reveal the size of transcripts and potentially the number of transcripts that correspond to the probe sequence. Quantitative real-time RT-PCR is the most commonly used technique for validation of microarray results, due to the enhanced level of sensitivity (Mackay et al. 2002). A review of the literature illustrates that microarray data generally are in good agreement with but not always confirmed by qRT-PCR (Jason et al. 2008; Linton et al. 2008). In the developmental study of watermelon, Wechter et al. (2008) reported that 72 of the 750 (9.6%) tissue-type quantitative-reactions were in conflict with the microarray results, thus 90.4% were in agreement. The failure to validate microarray data with qRT-PCR is often apportioned to the use of tissue sources which are, inadvertently, at different developmental stages (Gregersen et al. 2005; Lee et al. 2008).

Hansen et al. (2009) pointed out one more possible error in qRT-PCR validation experiments, especially when the research focuses on gene families. The study was conducted with primers designed to homologous regions within selected gene families. The rationale was to capture the 'average' expression level within a gene family; hence the qRT-PCR results were compared to an average of the microarray data which combined the expression of the alleles. Adopting such an approach resulted in a good correlation between the microarray and the qRT-PCR results, although the allelic variation observed and reported as part of the study (Hansen et al. 2009) was lost. Extrapolating from this rationale, we would like to urge caution when designing primers for qRT-PCR validation, because to design primers requires full knowledge of the allelic complement in any given genome. Without full sequence information used for primer design, gene expression observed using qRT-PCR can be an over- or under-estimation of relevant gene expression, thus compromising the attempt to get valid microarray derived results.

Identifying the functions of differentially expressed genes may be considered the ultimate validation of microarray data. Though the experiments may be carried out some time later, each level of data validation (mRNA, protein and

function) should be considered at the microarray experimental design stage, to allow the necessary additional controlled samples to be obtained (White and Salamonsen 2005).

17.3.5 Additional Regulatory Mechanisms in Gene Expression

The limitations of DNA microarray analysis discussed so far create challenges that require creative advances both in the technology and its application. However, a fundamental criticism of the DNA microarray methodology concerns the relevance of mRNA expression levels in establishing the actual functional levels of proteins. The transcription of genomic DNA to mRNA is only one of many stages involved in the biosynthesis of functional proteins. Alternative splicing, translational control, post-translational modification and the sub-cellular localisation of proteins are all additional potential regulatory stages that can control the production of functional proteins. It is clear that mRNA levels may not necessarily correlate with protein activity. While in some applications the product of the mRNA is not a concern for most experiments However, in most research-orientated applications the aim is to identify changes in mRNA profiles that are functional, and impact on protein expression (Walker and Hughes 2008).

17.3.6 Applications of DNA Microarrays

DNA microarray technology has revolutionized life-science research. Using arrays, researchers can examine the full complexity of a genome in a single experiment, allowing them to identify and study complex genetic regulatory networks and to begin to understand biology on a genome-wide scale. Arrays have been applied to studies in gene expression, genome mapping, SNP discrimination, transcription factor activity, toxicity, pathogen identification and detection, and many other applications. Walker and Hughes (2008) suggested the term: “Delineating molecular pathways” when analysing microarray results. Experimentation using DNA microarrays requires no previous knowledge about which genes might be of interest since expression from very large sets of genes can be examined in an unbiased manner. The technology can be used to identify unexpected transcriptional targets downstream of any stage in a molecular pathway, if those stages can either be specifically activated or repressed experimentally. This was demonstrated during the study the effect of antisense suppression of the C-hordein coding gene in barley (see below; Hansen et al. 2007).

Of particular relevance to this discussion is the utility of microarray as a platform to assist the development of functional markers which are applied as within the emerging concept genomic-assisted breeding. In next section we describe progress towards this objective.

17.4 Specific Example: Studies of Profiling Global Gene Expression During Grain Filling in Monocot

In the preceding sections we introduced the concept of genomic-assisted breeding, and the relevance of functional markers, following this we discussed the emerging interest in developing microarray and its associated output as a tool to support the identification of functional marker. In this section the objective is to illustrate with examples the utility of microarray as a tool to assist genomic-assisted breeding for quality traits.

17.4.1 Microarray in Plants: Specific Focus on Grain Fillings Studies in Monocots

Over the past decade, advances in genomic technologies have resulted in a variety of microarray platforms that can be used for global gene expression profiling. Microarrays only measure the steady state levels of those genes for which a probe, either a clone or sequence, is available and, therefore, should be considered a targeted approach to expression profiling. The application of microarray for studying grain development in monocots continues to increase. Table 17.2 provides a snapshot of the latest publications describing crop, type of method used, number of genes spotted and how many genes were shown to be significantly regulated during the period of study.

A common feature of the studies cited above is that most of the experimental plant material was grown under controlled conditions either in greenhouses or growth cabinets. Given the fact that a ‘systems approach’ must integrate the impact of the environment and since environment has a significant impact on plant performance, extrapolation of results from glasshouse-grown material to field-grown material is not straightforward. Recent studies have demonstrated the general utility of microarray analysis of field-grown plants (Duan and Sun 2005; Lu et al. 2005). Moreover, when comparing plant material grown in controlled conditions with field-grown material significant differences have been illustrated (Dhanaraj et al. 2007).

17.4.2 Large Arrays vs Tissue/Pathway Specific Approaches

The principal benefit of microarray technology is that it can simultaneously measure the differential expression of a large numbers of genes and thereby identify the relationship between genes involved in the control and regulation of cellular development and response of any organism. As a result of the volume of data generated and the type of analysis applied, typically microarray results are presented as a long list of genes judged to be significantly regulated. Table 17.2 shows the number of

Table 17.2 Examples of microarray based grain development studies: studies on grain filling of rice, wheat and barley using various global expression techniques

Crop	Type of global analysis	Source and number of genes spotted/ investigated	Description of targets	Number of differentially expressed genes (P < 0.05)
Rice[a]	Affymetrix	21,000	33 Rice samples, including 17 from	269
		GeneChip of sequenced rice genome	various stages of grain filling	1.3%
Rice[b, j]	Affymetrix	22,000	Wt vs. endospermless mutant at 2–3, 4–6	1,600
		Custom oligo DNA array	DAP	7.2%
Wheat[c]	cDNA chip	7,835	Wheat endosperm of 3, 7, 14, 21, 28	2,237
		Collected EST's from various libraries	and 35 DAF	28.6%
Wheat[d]	cDNA chip	2,304	Developing grain at 5–40 DAP	326
		Grain specific cDNA		14.1%
Wheat[e]	Affymetrix	55,052	Developing caryopses at	14,550
		Wheat genome array	6, 8, 10, 12, 14, 17, 21, 28, 35, 42 DAA	26.4%
Barley[f]	Differential display	11,787	Pericarp (0, 6, 12, 16 DAF), embryo (12,	2,384
		Seed cDNA libraries	18, 26 DAF), endosperm (0, 6, 12, 16, 26 DAF)	20.2%
Barley[g]	Differential display	1,421	Developing grain at 0, 2, 4, 6, 8, 10, 12	337
		Developing caryopses, etiolated seedlings and roots	DAF	23%
Barley[h]	Differential display	620	Developing grain at 0, 1, 2, 3, 4, 5, 6, 7	61
		Developing caryopses, etiolated seedlings and roots	DAF	9.8%

(continued)

Table 17.2 (continued)

Crop	Type of global analysis	Source and number of genes spotted/investigated	Description of targets	Number of differentially expressed genes (P < 0.05)
Barley[i, k]	cDNA chip	1,035	Developing grain at 10, 15, 18, 20, 25	501
		Pathway enriched grain specific cDNA	DAP	48.4%

[a] Zhu et al. (2003)
[b] Kondou et al. (2006)
[c] Laudencia-Chingcuanco et al. (2007)
[d] Laudencia-Chingcuanco et al. (2006)
[e] Wan et al. (2008)
[f] Sreenivasulu et al. (2006)
[g] Sreenivasulu et al. (2004)
[h] Sreenivasulu et al. (2002)
[i] Hansen et al. (2009)
[j] Material was grown in pots outdoors
[k] Material derived from field-grown plants
Days after flowering (DAF), days after pollination (DAP) and days after anthesis (DAA)

genes thought to be to expressed differentially among different monocots during grain developments (the list of genes are available in the cited articles). Whilst generating list of genes can have merit, there is an increasing realisation that combining molecular data derived from microarrays with information related to metabolic pathways is more useful. This combination provides a powerful tool to unravel metabolic control, regulation and cellular response to a range of stimuli, thereby offer greater insight into an organism's phenotype. The ability to link microarray data to biochemical data is being fostered by advances in bioinformatics: the application of Artificial Neural Networks as a result pathway specific analysis (also know as functional enrichment) is becoming popular (Curtis et al. 2005). However, despite the progress, achieving the ambition is not trivial even with well-annotated model organisms (human, mouse, or rat). This is not due to a lack of information, for the identified genes, but the magnitude of the task related collecting and examining gene information from multiple databases for thousands of genes (Olson 2006). For those working with plant systems, the matter is further compounded by the lack of fully annotated genome data for the major cereals. To date rice is the only cereal genome comprehensively annotated. Although the DNA sequences for cvs. Nipponbare and 93-11 are complete, rice genome sequence resources are constantly being revised and gene annotation updated (Collard et al. 2008). Therefore in the light of the cereal genomics it is reasonable to question the efficacy of microarrays fabricated from many thousands of cDNA or oligos, where a significant percentage of the genes are not fully annotated.

Instead of using large-scale microarrays for which data analysis may contribute to a loss of relevant information, a possible alternative is to employ organ/tissue/cell specific arrays that contain lower number of genes that are analysed in a pathway specific manner dictated by the relevant biological question. The viability and usefulness of this approach is illustrated in the next section, presenting the results of a grain specific microarray combined with pathway specific analysis.

17.4.3 Can Focused Microarray Follow Predicted Changes During Grain Development?

When perturbing and modulating plant regulatory mechanisms or metabolic pathways by genetic transformation, pleiotropic effects will invariably follow as interconnected pathways will also be affected (Brown 2002; Gupta and Kaur 2005; Hansen et al. 2007; Lange et al. 2007). Under these conditions microarray analysis could be a very powerful tool to elucidate and evaluate changes in metabolism that underpin the phenotype of the transgenic lines. Hansen et al. (2007) reported the use a grain-specific microarray to study the changes in the transcriptome of a transgenic barley line containing an antisense C-hordein gene. The C-hordein antisense line is characterised by marked changes in the targeted storage protein and as consequence in the overall amino acid profile (Lange et al. 2007). The reported custom-made cDNA array for developing grains of barley contains a comprehensive set of genes known to be involved in nitrogen mobilization, transport and AA metabolism. The results of the transcriptome analysis indicated that the transcript encoding C-hordein was down-regulated and this showed excellent correlation with the relative changes in the proportion of storage proteins and amino acid composition in the grain. As an apparent consequence, the level of sulphur-rich B/γ- and D-hordeins also increased in the transgenic line (Lange et al. 2007), which correlated with a higher steady state mRNA level of the corresponding genes. The increased synthesis of sulphur-rich hordeins appeared to increase the demand for sulphur-rich amino acids (cysteine and methionine), resulting in an up-regulation of key genes in the appropriate biosynthetic pathways. This study demonstrated the utility of the grain-specific cDNA microarray analysis to detect perturbations induced by an antisense suppression of plant processes (Hansen et al. 2007).

17.4.4 Extension of the Study to Field Grown Material: Expression of Alleles Coding Storage Proteins During Grain Development

The microarray analysis is expected to provide important information on the general transcriptional regulation of basic metabolism in the developing seeds and lead to the identification of genes that show significant up- or down-regulation in

response to altered genetic makeup or environmental effects. The grain-specific cDNA microarray reported by Hansen et al. (2007) was used in a further study designed to investigate amino acid biosynthesis and storage protein accumulation in the developing grains of field-grown barley (Hansen et al. 2009). The aim was to analyse the transcription of genes associated with the biosynthesis of storage products during the development of field grown barley grains and identify co-regulated genes. A distance matrix was constructed which enabled the identification of three clusters corresponding to the early, middle and late phase of grain development. The gene expression pattern associated with the clusters was investigated using a pathway-specific analysis with specific reference to the temporal expression levels of a range of genes involved mainly in the photosynthesis process, amino acid and storage protein metabolism. Hansen et al. (2009) conclude that the grain-specific microarray is a reliable and cost-effective tool for monitoring temporal changes in the transcriptome of the major metabolic pathways in the barley grain. Moreover, this microarray is sensitive enough to monitor differences in the gene expression profiles of different homologues from the storage protein families using field-grown material.

17.4.5 From Microarray to In Silico Plant Design

Throughout this chapter the aim has been to develop an argument that supports DNA microarray as having a pivotal role within genomic-assisted breeding. While this concept has yet to emerge fully, it is interesting to speculate further and conceptualise a vision for future plant breeding. We present a hypothetical model of ‘virtual genomic-assisted breeding’ (Fig. 17.2). Our model combines elements of

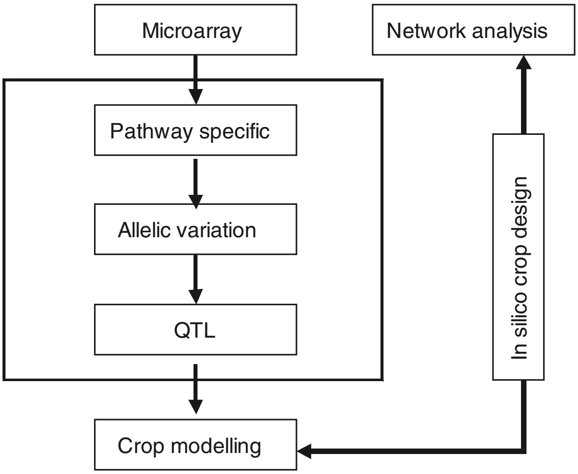


Fig. 17.2 The combined use of molecular genomic based tools with more traditional plant breeding tools supports genomic-assisted breeding. However, as further advances are made we propose that ultimately predictive in silico crop design will be possible

the emerging genomic-assisted breeding with crop modelling and network analysis interfaced with *in-silico* crop design.

Crop modelling has already demonstrated considerable power in predicting crop performance based on plant physiology and agronomic data (Hammer et al. 2002). Yin et al. (2004) suggested the potential synergy between crop modelling and the emerging molecular genetics as a way of strengthen the crop models, and more specifically, bridge the gap between plant physiologist and geneticists.

Similarly network analysis, which has already been applied to biochemical pathways and QTL analysis, as discussed above, has the potential to interface between the pathway specific analysis, which forms part of the 'bottom-up' system analysis and the crop modelling, which is a 'top-down' system analysis (Hammer et al. 2004).

17.5 Conclusion

Future crop development requires the successful integration of molecular tools with traditional breeding techniques. In the preceding we have outlined the need for a genomic-assisted breeding approach improve nutritional quality of barley. More specifically, the potential of microarray as a tool to support breeding for crop quality traits was evaluated and discussed. While microarray has received much attention, the true potential has yet to be realised, in part due the many technical issues surrounding the technique. In the light of limited success, we propose that using smaller but focused grain-specific microarray coupled with pathway-specific analysis offers a reliable tool for monitoring temporal changes in the transcriptome of the major metabolic pathways in the barley grain. Such an approach is able to monitor differences in the gene expression profiles of storage protein homologues. Therefore, microarray analysis could provide the knowledge required to design an improved AA profile with the intriguing possibility of breeding selectively for specific alleles/homologues to confer enhanced amino acid profile of the barley storage proteins (Hansen et al. 2009).

References

Alba R, Fei Z, Paxton Payton P, Liu Y, Moore SL, Debbie P, Cohn J, D'Ascenzo M, Gordon JS, Rose JKC, Martin G, Tanksley SD, Bouzayen M, Molly M, Jahn MM, Giovannoni J (2004) ESTs, cDNA microarrays, and gene expression profiling: tools for dissecting plant physiology and development. Plant J 39:697–714

Alwine JC, Kemp DJ, Stark GR (1977) Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and *hyb*ridization with DNA probes. Proc Natl Acad Sci USA 74:5350–5354

Andersen JR, Lubberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554–560

Baik B-K, Ullrich SE (2008) Barley for food: characteristics, improvement, and renewed interest. J Cereal Sci 48:233–242

Bao J-Y, Lee S, Chen C, Zhang X-Q, Zhang Y, Liu S-Q, Clark T, Wang J, Cao M-L, Yang H-M, Wang SM, Yu J (2005) Serial analysis of gene expression study of a hybrid rice strain (LYP9) and its parental cultivars. Plant Physiol 138:1216–1231

Bethke PC, Jakobsen JV, Jones RL (2000) Barley biotechnology. In: Black M, Bewley JD (eds) Seed technology and its biological basis. CRC Press LLC, Sheffield, pp 184–225
Boisen S, Hvelplund T, Weisbjerg MR (2000) Ideal amino acid profiles as a basis for feed protein evaluation. Livestock Prod Sci 64:239–251
Breitling R (2006) Biological microarray interpretation: the rules of engagement. Biochim Biophys Acta 1759:319–327
Breitling R, Herzyk P (2005) Rank-based methods as a non-parametric alternative of the T-statistic for the analysis of biological microarraydata. J Bioinform Comput Biol 3: 1171–1189
Breitling RP, Armengaud P, Amtmann A, Herzyk P (2004) Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett 573:83–92
Brem RB,Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. Science 296:752–775
Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J, Corcoran K (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol 6:630–634
Brown JKM (2002) Yield penalties of disease resistance in crops. Curr Opin Plant Biol 5:1–6
Brown AHD, Zohary D, Nevo E (1978) Outcrossing rates and heterozygosity in natural populations of *Hordeum spontaneum* Koch in Israel. Heredity 41:49–62
Chen JJ, Delongchamp RR, Tsai C-A, Hsueh H-M, Sistare F, Thompson KL, Desai VG, Fuscoe JC (2004) Analysis of variance components in gene expression data. Bioinformatics 20:1436–1446
Churchill GA (2002) Fundamentals of experimental design for cDNA microarrays. Nat Genet 32 Suppl:490–495
Collard BCY, Cruz CMV, McNally KL, Virk PS, Mackill DJ (2008) Rice molecular breeding laboratories in the genomics era: current status and future considerations. Int J Plant Genomics 2008: Article ID 524847, 25 pages
Cui X, Churchill GA (2003) Statistical tests for differential expression in cDNA microarray experiments. Genome Biol 4:210
Curtis RK, Oresic M, Vidal-Puig A (2005) Pathways to the analysis of microarray data. Trends Biotechnol 23:429–435
Dhanaraj AL, Alkharouf NW, Beard HS, Chouikha IB, Matthews BF, Wei H, Arora R, Rowland LJ (2007) Major differences observed in transcript profiles of blueberry during cold acclimation under field and cold room conditions. Planta 225:735–751
Diatchenko L, Lau Y-FC, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov E, Siebert PD (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc Natl Acad Sci USA 93:6025–6030
Doll H (1980) A nearly non-functional mutant allele of the storage protein locus Hor2 in barley. Hereditas 93:217–222
Duan MJ, Sun SSM (2005) Profiling the expression of genes controlling rice grain quality. Plant Mol Physiol 59:165–178
Dudoit S, Yang YH, Callow MJ, Speed TP (2002) Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. Stat Sin 12:111–139
Food and Drug Administration HHS (2008) Food labeling: health claims; soluble fiber from certain foods and risk of coronary heart disease. Final rule. Fed Regist 73:47828–47829
Galili G, Amir R, Hoefgen R, Hesse H (2005) Improving the levels of essential amino acids and sulfur metabolites in plants. Biol Chem 386:817–831
Gregersen PL, Brinch-Pedersen H, Holm PB (2005) A Microarray-based comparative analysis of gene expression profiles during grain development in transgenic and wild type wheat. Trans Res 14:887–905

Guo M, Rupe MA, Yang X, Crasta O, Zinselmeier C, Smith OS, Bowen B (2006) Genome-wide transcript analysis of maize hybrids allelic additive gene expression and yield heterosis. Theor Appl Genet 113:831–845

Gupta AK, Kaur N (2005) Sugar signalling and gene expression in relation to carbohydrate metabolism under abiotic stresses in plants. J Biosci 30:761–776

Hammer GL, Kropff MJ, Sinclair TR, Porter JR (2002) Future contributions of crop modelling: from heuristics and supporting decision making to understanding genetic regulation and aiding crop improvement. Eur J Agron 18:15–31

Hammer GL, Sinclair TR, Chapman SC, van Oosterom E (2004) On systems thinking, systems biology, and the in silico plant. Plant Physiol 134:909–911

Han B, Altman NS, Mong JA, Klein LC, Pfaff DW, Vandenbergh DJ (2008) Comparing quantitative trait loci and gene expression data. Adv Bioinform 2008: Article ID 719818, 6 pages doi:10.1155/2008/719818

Hansen M, Lange M, Friis C, Dionisio G, Holm, PB, Vincze E (2007) Antisense mediated suppression of C-hordein biosynthesis in the barley grain results in correlated changes in the transcriptome, protein profile and amino acid composition. J Exp Bot 58:3987–3995

Hansen M, Bowra S, Friis C, Holm PB, Vincze E (2009) A pathway specific microarray analysis highlights the complex and coordinated transcriptional networks of the developing grain of field grown barley. J Exp Bot 60:153–167

Harlan JR, Zohary D (1966) Distribution of wild wheats and barley. Science 153:1074–1079

Hirel B, Gouis JL, Ney B, Gallais A (2007) The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. J Exp Bot 58:2369–2387

Hospital F (2008) Challenges for effective marker-assisted selection in plants. Genetica doi:10.1007/s10709-008-9307-1

Huang SS, Adams WR, Zhou Q, Malloy KP, Voyles DA, Anthony J, Kriz AL, Luethy MH (2004) Improving nutritional quality of maize proteins by expressing sense and antisense zein genes. J Agri Food Chem 52:1958–1964

Jason A, Holliday SG, White RR, Bohlmann J, Aitken SN (2008) Global monitoring of autumn gene expression within and among phenotypically divergent populations of Sitka spruce (*Picea sitchensis*). New Phytol 178:103–122

Kirk RE (1996) Practical significance: a concept whose time has come. Educ Psychol Meas 56:746–759

Kohno-Murase J, Murase M, Ichikawa H, Imamura J (1995) Improvement in the quality of seed storage protein by transformation of *Brassica napus* with an antisense gene for cruciferin. Theor Appl Genet 91:627–631

Kondou H, Ooka H, Yamada H, Satoh K, Kikuchi S, Takahara Y, Yamamoto K (2006) Microarray analysis of gene expression at initial stages of rice seed development. Breed Sci 56:235–242

Kothapalli R, Yoder SJ, Mane S, Loughran Jr TP (2002) Microarray results: how accurate are they? BMC Bioinform 3:22

Kreis M, Shewry PR, Forde BG, Rahman S, Miflin BJ (1983) Molecular analysis of a mutation conferring the high lysine phenotype on the grain of barley (*Hordeum-vulgare*). Cell 34:161–167

Kreis M, Shewry PR, Forde BG, Rahman S, Bahramian MB, Miflin BJ (1984) Molecular analysis of the effects of the *Lys 3a* gene on the expression of Hor loci in developing endosperms of barley (*Hordeum-vulgare*-L). Biochem Genet 22:231–255

Lange M, Vincze E, Wieser H, Schjørring JK, Holm PB (2007) Effect of an antisense C-hordein gene on the storage protein composition in the barley seed. J Agri Food Chem 55:6074–6081

Laudencia-Chingcuanco DL, Stamova BS, You FM, Lazo GR, Cui X, Anderson OD (2006) Analysis of the wheat endosperm transcriptome. J Appl Genet 47:287–302

Laudencia-Chingcuanco DL, Stamova BS, You FM, Lazo GR, Beckles DM, Anderson OD (2007) Transcriptional profiling of wheat caryopsis development using cDNA microarrays. Plant Mol Biol 63:651–668

Lasztity R (ed) (1996) Barley proteins. In: The chemistry of cereals proteins. CRC Press, Boca Raton, FL, pp 159–183
Lee M (1995) DNA markers and plant breeding programs. Adv Agron 55:265–344
Lee MH, Kim B, Song SK, Heo JO, Yu NI, Lee SA, Kim M, Kim DG, Sohn SO, Lim CE, Chang KS, Lee MM (2008) Large-scale analysis of the GRAS gene family in *Arabidopsis thaliana*. Plant Mol Biol 67:659–670
Lee M-LT, Frank C, Kuo FC, Whitmorei GA, Sklar J (2000) Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. Proc Natl Acad Sci USA 97:9834–9839
Liang P, Pardee AB (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 257:967–971
Linton KM, Hey Y, Saunders E, Jeziorska M, Denton J, Wilson CL, Swindell R, Dibben S, Miller CJ, Pepper SD, Radford JA, Freemont AJ (2008) Acquisition of biologically relevant gene expression data by Affymetrix microarray analysis of archival formalin-fixed paraffin-embedded tumours. Br J Cancer 98:1403–1414
Lockhart DJ, Winzeler EA (2000) Genomics, gene expression and DNA arrays. Nature 405:827–836
Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol 14:1675–1680
Lu C, Hawkesford MJ, Barraclough BJ, Poulton PR, Wilson ID, Barker GL, Edwards KJ (2005) Markedly difference gene expression in wheat grown with organic and inorganic fertilizer. Proc Res Soc B 272:1901–1908
Mackay IM, Arden KE, Nitsche A (2002) Real time PCR in virology. Nucleic Acids Res 30:1292–1305
Maruta Y, Ueki J, Saito H, Nitta N, Imaseki H (2002) Transgenic rice with reduced glutelin content by transformation with glutelin A antisense gene. Mol Breed 8:273–284
Morandini P, Salamini F (2003) Plant biotechnology and breeding: allied for years to come. Trends Plant Sci 8:70–75
Morrell PL, Clegg MT (2007) Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. Proc Natl Acad Sci USA 104:3289–3294
Munck L (1992) The case of high-lysine barley breeding. In: Shewry PR (ed) Barley: genetics, biochemistry, molecular biology, biotechnology. C.A.B. International, Wallingford, pp 573–601
Newman CW, Newman RK (2006) A brief history of barley food. Cereal Food World 51:4–7
Nolan T, Hands RE, Bustin SA (2006) Quantification of mRNA using real-time RT-PCR. Nat Protoc 1:1559–1582
Olson NE (2006) The microarray data analysis process: from raw data to biological significance. NeuroRx 3:373–383
Panfili G, Fratianni F, Criscio TD, Marconi E (2008) Tocol and β-glucan levels in barley varieties and in pearling by-products. Food Chem 107:84–91
Parry MAJ, Flexas J, Medrano H (2005) Prospects for crop production under drought: research priorities and future directions. Ann Appl Biol 147:211–226
Parry MAJ, Madgwick PJ, Carvahlo JFC, Andralojc PJ (2007) Prospects for increasing photosynthesis by overcoming the limitations of Rubisco. J Agri Sci 145:31–43
Pastinen T, Hudson TJ (2004) Cis-acting regulatory variation in the human genome. Science 306:647–650
Pastinen T, Sladek R, Gurd S, Sammak A, Ge B, Lepage P, Lavergne K, Villeneuve A, Gaudin T, Brändström H, Beck A, Verner A, Kingsley J, Harmsen E, Labuda D, Morgan K, Vohl MC, Naumova AK, Sinnett D, Hudson TJ (2004) A survey of genetic and epigenetic variation affecting human gene expression. Physiol Genomics 16:184–193
Peleman JD, van der Voort JR (2003) Breeding by design. Trends Plant Sci 8:330–334
Pins JJ, Kaur H (2006) A review of the effects of barley b-glucan on cardiovascular and diabetic risk. Cereal Foods World 51:8–11

Powlson DS, Riche AB, Shield I (2005) Biofuels and other approaches for decreasing fossil fuel emissions from agriculture. Ann Appl Biol 146:193–201

Ribaut J-M, Hoisington D (1998) Marker-assisted selection: new tools and strategies. Trends Plant Sci 3: 236–239

Rieseberg LH, Widmer A, Arntz AM, Burke JM, (2003) The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. Philos Trans R Soc Lond B Biol Sci 358:1141–1147

Rockett JC, Hellmann GM (2004) Confirming microarray data - is it really necessary? Genomics 83:541–549

Saisho D, Purugganan MD (2007) Molecular phylogeography of domesticated barley traces expansion of agriculture in the Old World. Genetics 177:1765–1776

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470

Segal G, Song RT, Messing J (2003) A new opaque variant of maize by a single dominant RNA-interference-inducing transgene. Genetics 165:387–397

Smyth GK, Speed T (2003) Normalization of cDNA microarray data. Methods 31:265–273

Sparrow DHB (1970) Some genetical aspects of malting quality. In: Nilan RA (ed) Barley genetics 2, proceedings of 2nd international barley genetics symposium, Washington State University Press, Pullman, pp 559–574

Sreenivasulu N, Altschmied L, Panitz R, Hahnel U, Michalek W, Weschke W, Wobus U (2002) Identification of genes specifically expressed in maternal and filial tissues of barley caryopses: a cDNA array analysis. Mol Genet Genomics 266:758–767

Sreenivasulu N, Altschmeid L, Radchuk V, Gubatz S, Wobus U, Weschke W (2004) Transcript profiles and deduced changes of metabolic pathways in maternal and filial tissues of developing barley grains. Plant J 37:539–553

Sreenivasulu N, Radchuk V, Strickert M, Miersch O, Weschke W, Wobus U (2006) Gene expression patterns reveal tissue-specific signalling networks controlling programmed cell death and ABA-regulated maturation in developing barley seeds. Plant J 47:310–327

Stupar RM, Hermanson PJ, Springer NM (2007) Nonadditive expression and parent-of-origin effects identified by microarray and allele-specific expression profiling of maize endosperm. Plant Physiol 145:411–425

Stupar RM, Gardiner JM, Oldre AG, Haun WJ, Chandler VL, Springer NM (2008) Gene expression analyses in maize inbreds and hybrids with varying levels of heterosis. BMC Plant Biol 2008 8:33

Tang T, François N, Glatigny A, Agier N, Mucchielli MH, Aggerbeck L, Delacroix H (2007) Expression ratio evaluation in two-colour microarray experiments is significantly improved by correcting image misalignment. Bioinformatics 23:2686–2691

Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 98:5116–5121

Varshney RK, Graner A, Sorrells, ME (2005a) Genomics-assisted breeding for crop improvement. Trends Plant Sci 10:621–630

Varshney RK, Graner A, Sorrells, ME (2005b) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55

Varshney RK, Langridge P, Graner A (2007) Application of genomics to molecular breeding of wheat and barley. Adv Genetics 58:121–155

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science 270:484–487

Vincze E, Bowra S (2005) Northerns revisited: a protocol that eliminates formaldehyde from the gel while enhancing resolution and sensitivity. Anal Biochem 342:356–357

Walker MS, Hughes TA (2008) Messenger RNA expression profiling using DNA microarray technology: diagnostic tool, scientific analysis or un-interpretable data? Int J Mol Med 21:13–17

Wan Y, Poole RL, Huttly AK, Toscano-Underwood C, Feeney K, Welham S, Gooding MJ, Mills C, Edwards KJ, Shewry PR, Mitchell RAC (2008) Transcriptome analysis of grain development in hexaploid wheat. BMC Genomics 2008 9:121

Wang X, Le Roy I, Nicodeme E, Li R, Wagner R, Petros C, Churchill GA, Harris S, Darvasi A, Kirilovsky J et al (2003) Using advanced intercross lines for high-resolution mapping of HDL cholesterol quantitative trait loci. Genome Res 13:1654–1664

Wang Z, Ni Z, Wu H, Nie X, Sun Q (2006) Heterosis in root development and differential gene expression between hybrids and their parental inbreds in wheat (*Triticum aestivum* L.). Theor Appl Genet 113:1283–1294

Wayne ML, McIntyre LM (2002) Combining mapping and arraying: an approach to candidate gene identification. Proc Natl Acad Sci USA **99**:14903–14906

Wechter WP, Levi A, Harris KR, Davis AR, Fei Z, Katzir N, Giovannoni JJ, Salman-Minkov A, Hernandez A, Thimmapuram J, Tadmor Y, Portnoy V, Trebitsh T (2008) Gene expression in developing watermelon fruit. BMC Genomics 9:275

White CA, Salamonsen LA (2005) A guide to issues in microarray analysis: application to endometrial biology. Reproduction 130:1–13

Wu LM, Ni ZF, Meng FR, Lin Z, Sun QX (2003) Cloning and characterization of leaf cDNAs that are differentially expressed between wheat hybrids and their parents. Mol Genet Genomics 270:281–286

Yin X, Struik PC, Kropff MJ (2004) Role of crop physiology in predicting gene-to-phenotype relationships. Trends Plant Sci 9:426–432

Yu J, Mohammad I, Othman MI, Farjo R, Zareparsi S, MacNee SP, Yoshida S, Swaroop A (2002) Evaluation and optimization of procedures for target labeling and hybridization of cDNA microarrays. Mol Vision 8:130–137

Zhang X, Richards EJ, Borevitz JO (2007) Genetic and epigenetic dissection of cis regulatory variation. Curr Opin Plant Biol 10:142–148

Zhu T, Budworth P, Chen W, Provart N, Chang H-S, Guimil S, Su W, Estes B, Zou G, Wang X (2003) Transcriptional control of nutrient partitioning during ricegrain filling. Plant Biotechnol J 1:59–70

Zohary D, Hopf M (2000) Domestication of plants in the old world: the origin and spread of cultivated plants in West Asia, Europe, and the Nile Valley. Oxford University Press, New York

Chapter 18
Unravelling Gene Function Through Mutagenesis

Andrea Hricová, Pedro Robles, and Víctor Quesada

Abstract In recent years, plant research has received a significant boost, partly due to the huge amount of data derived from the sequencing projects of some genomes, especially those obtained in the model system *Arabidopsis thaliana*. The use of this crucifer in many plant biology laboratories all over the world, together with the development of new technologies, have considerably increased research resources for dissecting gene function. Forward and reverse genetic approaches are still fundamental for assessing gene function through the isolation and characterization of mutants. In this chapter, we first take an overview, using Arabidopsis as a reference, of the methodologies and procedures used to infer gene function in plants through mutagenesis screens and look at how high-throughput analysis of gene-indexed mutant collections is contributing to assessing plant gene function on a genomic scale. We then discuss the forward and reverse genetic approaches currently used in crops for understanding gene function, highlighting the importance of mutagenesis as a basis for generating useful genetic variants and its significance for crop improvement.

18.1 Introduction

The plant biology community entered the "genomic" era when the genomic sequences of Arabidopsis and rice were obtained in 2000 and 2002, respectively (The Arabidopsis Genome Initiative [AGI] 2000; Goff et al. 2002; Yu et al. 2002). Since then, the greatest challenge for plant research has become to assign a function to all the genes identified in the already sequenced genomes and in those whose sequencing

A. Hricová (✉)
Institute of Plant Genetics and Biotechnology, Slovak Academy of Sciences, Nitra, Slovak Republic
e-mail: andrea.hricova@savba.sk

P. Robles and V. Quesada
División de Genética and Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, Alicante, Spain

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_18, © Springer Science+Business Media B.V. 2010

is near completion. To this end, different high-throughput genomic technologies such as microarrays or shotgun mass spectrometry (Brown and Botstein 1999; Yates 2004; Steinmetz and Davis 2004) have been developed and successfully applied in Arabidopsis. Nonetheless, in spite of the vast amount of data generated by these methodologies, in most cases they are not sufficient to unequivocally assign gene function (Alonso and Ecker 2006).

As an alternative or complementary way for dissecting gene function, classical forward genetics based on mutant analysis has proved decisive in the identification of genes controlling different aspects of plant biology. The availability of whole-genome sequences has made it possible to develop high-throughput mapping methods that have significantly improved gene cloning (Jander et al. 2002; Peters et al. 2004; Ponce et al. 2006). This, together with the constantly growing collections of gene-tagged (or gene-indexed) mutants, the development of TILLING and RNAi-mediated gene silencing strategies, is making possible to carry out reverse genetics in plants. In this review, we focus on how these forward and reverse genetics approaches are currently being used in Arabidopsis and crops on a genomic scale to ascertain gene function.

18.2 Forward Genetics

Arabidopsis thaliana is a model organism widely used in plant biology research to perform genetic and molecular analyses of complex biological processes (Meyerowitz et al. 1994). Arabidopsis has become an extremely useful tool for the functional analysis of genes through the isolation and characterization of thousands of mutants affected in different aspects of plant growth, development or physiology.

In classical forward genetic approaches, gene function is assigned based on the analysis of the mutant phenotype generated by a mutation perturbing the activity of a gene, compared with that of wild-type plants. In the course of the last two decades, a vast number of mutagenesis screens have been carried out in *Arabidopsis thaliana* in order to identify genes that underlie a specific developmental or physiological plant process (e.g. leaf [Berná et al. 1999], root [Benfey et al. 1993], flower [Bowman et al. 1989], seed [Mayer et al. 1991; Léon-Kloosterziel et al. 1996; McElver et al. 2001] or trichome cell [Hülskamp et al. 1994] development, tolerance to salt stress [Wu et al. 1996; Quesada et al. 2000] or response to vernalization [Chandler et al. 1996]). The rationale behind this approach is that the mutation perturbs the function of a gene required for a specific process and, as a result, causes an abnormal phenotype in the plant. For instance, if the mutant plant shows an alteration in leaf morphology it can be inferred that the gene perturbed is required for normal leaf development. Following this approach, hundreds of genes have been identified in Arabidopsis and other plant species by mutagenesis in the last 20 years and many of the genes underlying these mutants have been cloned. The main drawback of this approach is the functional redundancy that exists in the

Arabidopsis genome, probably as a consequence of an early event of whole-genome duplication (The AGI 2000). Accordingly, loss-of-function mutations in a particular gene might not result in a mutant phenotype clearly distinguishable from the wild-type one. In such cases, in order to obtain a mutant phenotype, it is necessary to inactivate (by mutation) two or more functionally related genes by constructing a double, triple or quadruple mutant (see for instance Eshed et al. 2001; Kumaran et al. 2002; Ditta et al. 2004; Adamczyk et al. 2007; Allen et al. 2007).

A wide range of chemical and physical mutagens has been used to obtain Arabidopsis mutants, differing in the extent (number of mutations inherited by the progeny) and the nature (ranging from the introduction of single nucleotide changes to large chromosomal rearrangements) of the damage caused to the genome. Irradiation with physical mutagens, such as fast neutrons, γ-rays, X-rays or carbon ions, has been extensively used in Arabidopsis, yielding a large number of mutants, which in some cases have been molecularly characterized (e.g. Redei and Koncz 1992; Shirley et al. 1992; Bruggemann et al. 1996; Cecchini et al. 1998; Niyogi et al. 1998; Shikazono et al. 1998, 2005; Quesada et al. 2000; Li and Zhang 2002). Ionizing radiations produce deletions, ranging from one to thousands nucleotides, as well as chromosomal rearrangements, typically resulting in loss-of-function or, in most cases, null alleles. The analysis of mutations induced by ionizing radiations has been instrumental in identifying genomic regions essential for gamete viability (Vizir and Mulligan 1999; Naito et al. 2005). However, since most large deletions are poorly transmitted to the progeny through the pollen (Naito et al. 2005), only a fraction of the ion-radiation-induced mutants are useful for further genetic and molecular analyses. Besides, identification of the deletion or reorganization in the genome is a very laborious task. Despite this, techniques such as subtractive hybridization (Straus and Ausbel 1990; Sun et al. 1992) and a new knockout methodology, Deleteagene (Li et al. 2001; Li and Zhang 2002) have been successfully applied for gene identification in deletion-mutants (Sun et al. 1992; Silverstone et al. 1998; Li et al. 2001).

As regards chemical agents, ethyl nitrosomethylurea (NMU) and ethyl methanesulphonate (EMS) are now the most extensively used mutagens in the Arabidopsis community when performing forward genetics. This is due to the high efficiency of these mutagens, which usually introduce hundreds of mutations per plant genome (Alonso and Ecker 2006). In addition, compared with irradiation-induced mutations, the EMS generates more embryonic and chlorophyll mutants and does not often produce double strand breaks leading to deletion or inversion mutations (Koorneef et al. 1982). NMU and EMS generally induce point mutations, mostly single-nucleotide substitutions, generating in most cases recessive alleles, although dominant or semi-dominant alleles have also been reported (see for instance, Berná et al. 1999; Soppe et al. 2000; Ochando et al. 2006; Würschum et al. 2006). Most EMS-derived alleles are hypomorphic and viable.

To identify mutants with a looked for phenotype from a mutagenesis screen, the first step is to mutagenize usually thousands of wild-type seeds (M_1 generation), sow them and led the M_1 plants growth to maturity. The M_2 seeds produced by the M_1 plants are collected, sowed, and the M_2 plants screened (Fig. 18.1a).

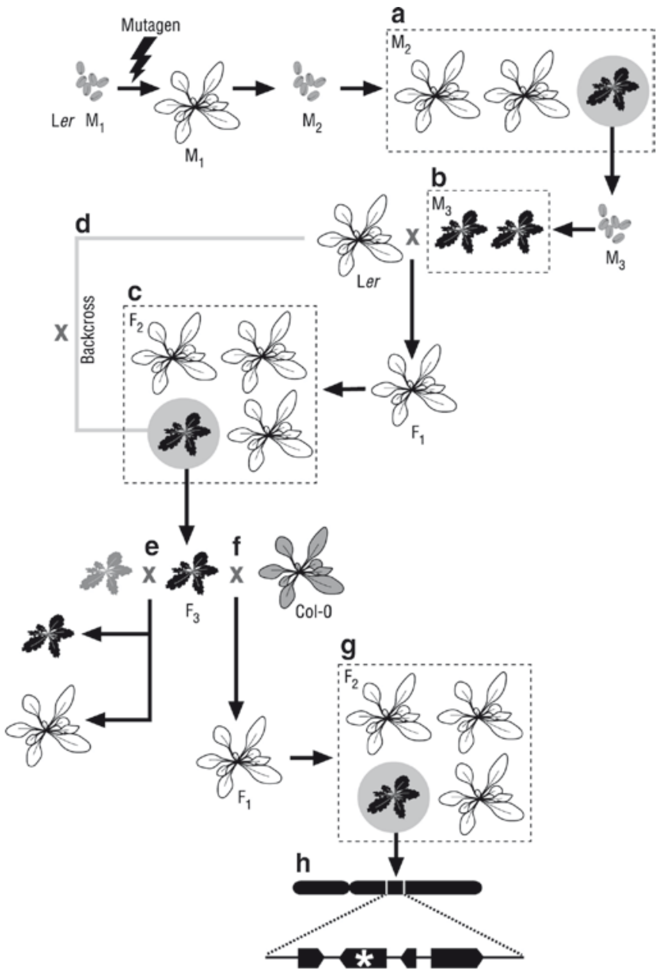


Fig. 18.1 Overview of the main steps followed in the forward genetic approach to establish gene function by mutagenesis. M_1 wild-type (e.g. L*er*) seeds are exposed to the mutagen to obtain M_2 progenies which are screened for mutant plants. (a) Isolation of M_2 mutant individuals with the desired phenotype and (b) cross of M_3 mutant plants with a hereditable phenotype to the wild type (e.g. L***er***) to (c) ascertain, after analysis of the F_1 and F_2 progenies, the inheritance pattern of the mutation. (d) The mutant is further backcrossed to the wild type (e.g. L***er***) to remove undesired additional mutations, (e) crossed to other mutants showing a similar phenotype for complementation analysis or (f) to a wild-type accession in a different genetic background (e.g. Col-0) for (g) gene mapping (by genotyping with polymorphic markers F_2 mutant plants) and (h) cloning. The asterisk denotes a mutation in a gene of a genomic candidate interval. Transcription units and intergenic regions are indicated by black boxes and lines, respectively. For simplicity, only the isolation of a recessive mutation affecting leaf development is shown

Although seeds of Arabidopsis contain around 15,000–20,000 cells (Jürgens 1994), only 2–3 of them contribute to the seed output. These cells are known as the number of genetically effective cells (Müller 1965; Li and Rédei 1969;

Grinikh and Shevchenko 1976; Rédei and Koncz 1992). The next step is to confirm the heritability of the mutant phenotype by sowing the M_3 seeds collected from the putative M_2 mutants and studying the phenotype of the M_3 plants (Fig. 18.1b). From each M_3 progeny with a fully heritable mutant phenotype, individuals are selected and back-crossed to the wild type, to determine the recessivity or dominance of the mutation in the F_1 progeny. To ascertain the monogenic nature of the mutation under study, F_2 inbred descendant from the F_1 plants are sown and studied (Fig. 18.1c). Given that the EMS and NMU cause numerous mutations in the genome, it is convenient to back-cross the mutant plants to the wild type at least three times and select the mutant plants in the resulting F_2 progenies in order to remove additional and undesired mutations (Fig. 18.1d). Comprehensive descriptions of methods used to carry out mutagenesis screens in Arabidopsis have been provided previously (Redei and Koncz 1992; Weigel and Glazebrook 2002; Østergaard and Yanofsky 2004; Kim et al. 2006).

Usually, the confirmed mutants are later classified into phenotypic classes in order to simplify allelism tests. The reason for this is the assumption that individuals exhibiting similar mutant phenotypes are probably carrying mutations affecting the same gene. Thus, if the F_1 progeny resulting from the cross of two different and recessive mutants grouped in the same phenotypic class shows a mutant phenotype, it is generally concluded that both mutations perturb the same gene. On the contrary, if the F_1 plants display a wild-type appearance, the mutations are probably within different genes (Fig. 18.1e). Eventually, the mutant plants can be crossed to previously described and characterized mutants with a similar phenotype in order to establish whether they are allelic and, therefore, affected in the same gene (Fig. 18.1e).

The ideal of a mutagenesis screen would be to identify all the genes involved in a particular biological process. For this purpose, it is necessary to isolate a sufficiently high number of mutants to obtain a collection of mutant alleles representative of all the genes responsible for a specific process. When this happens, the mutant screen has reached "saturation" and several mutant alleles are obtained for each gene (Jürgens et al. 1991; Berná et al. 1999). Saturation mutagenesis can be carried out only in model organisms such as Arabidopsis, where large progenies of mutagenized individuals can be screened. Given the high efficiency of chemical mutagens, it is possible to obtain alleles of different phenotypic strengths (an allelic series) for a gene of interest, contributing to a better understanding of its function in the process under study.

18.2.1 Gene Mapping and Cloning

Map-based cloning or positional cloning is the most frequently used method to identify a gene whose perturbation by a chemical or physical mutagen causes a mutant phenotype, when these mutations are untagged. For this purpose, the first step is to obtain a mapping population segregating for the mutant phenotype, by crossing the mutant of interest to a wild-type plant in a different genetic background

(Fig. 18.1f). Columbia-0 (Col-0) and Landsberg *erecta* (L*er*) are the most often used ecotypes in Arabidopsis laboratories, irrespectively of whether Col-0 is the mutant genetic background and L*er* that of the wild type or vice versa. If the mutation is recessive, as usually occurs with the EMS, the DNA of F_2 plants displaying the mutant phenotype (a quarter of the total number of F_2 plants if the penetrance of the mutation is complete) is individually extracted and used to carry out linkage analysis to polymorphic markers whose map positions are previously known (Fig. 18.1g). Many different collections are currently available, including thousands of polymorphic markers between Col-0 and L*er* such as the Cereon Arabidopsis Polymorphic Collection (Cambridge, MA, USA). Most markers in this collection are single nucleotide (SNPs) and insertion-deletion (InDel) polymorphisms.

In recent years, different high-throughput mapping approaches based on the polymerase chain reaction (PCR) have been developed in Arabidopsis for low resolution mapping (Ponce et al. 1999; Jander et al. 2002; Peters et al. 2004; Ponce et al. 2006). In addition, high-density DNA arrays have been designed for mapping deletions caused by physical mutagens (Spiegelman et al. 2000; Gong et al. 2004). To clone the gene of interest, the candidate genomic interval must be further narrowed-down by genotyping additional F_2 individuals (Fig. 18.1g, h). Given the large collection of molecular markers currently available for the Arabidopsis community, less than 2 months is usually enough to delimit a 100–200 kb candidate interval (Ponce et al. 2006). This makes final identification of the gene a straightforward task, which is in sharp contrast to the situation existing prior to the completion of the sequencing of the Arabidopsis genome (The AGI 2000), when cloning a gene of interest by chromosome-walking (Bender et al. 1983) used to take several years.

18.3 Reverse Genetics

Knowledge of the whole Arabidopsis genome sequence, combined with the generation of nearly saturation collections of randomly obtained T-DNA-tagged lines or indexed-mutants, has led to the development of reverse genetics and functional genomics in this model system. In less than a decade, a new scenario has been created in which, in a few weeks, it is possible to obtain a mutant allele of a particular gene.

18.3.1 Arabidopsis Genome Sequence

The almost complete Arabidopsis genome sequence was released at the end of 2000, becoming the first plant genome to be unmasked (The AGI 2000). This date opened up a new *omics* era for plant biologists, whose major goal became to assign a function to every one of the 25,498 predicted genes by the year 2010 (Chory et al.

2000; Somerville and Dangl 2000). Two years later, a high quality draft of a crop genome, rice, was published (Goff et al. 2002; Yu et al. 2002). To date Arabidopsis (Swarbreck et al. 2008) and rice (International Rice Genome Sequencing Project 2005) are the only plants whose genomes have been completely sequenced and annotated to a high standard.

Since 2000, first TIGR (Haas et al. 2005) and then TAIR (Swarbreck et al. 2008) have re-annotated the sequence of the Arabidopsis genome seven times thanks to new sequencing and array-based technologies. This has made possible to discover new, previously undetected genes. The most recent update of the Arabidopsis genome sequence and annotation (TAIR7, Swarbreck et al. 2008) has led to 120 out of 135 Mb of the genome sequence being completed, and 26,819 nuclear protein-coding genes being predicted, 3,866 of them alternatively spliced. Of the predicted protein-coding genes, 22,032 are transcriptionally supported by at least one cDNA and/or EST (Expressed Sequence Tag) and 4,000 show similarity to uncharacterized proteins. Moreover, 758 genes display no significant homology to GeneBank database (http://www.ncbi.nlm.nih.gov/Genbank/) accessions, of which 286 have no transcriptional support and probably correspond to erroneous gene predictions. Gene density is 4.44 kb per gene. Additionally, 1,054 nuclear RNA-coding genes and 3,889 pseudogenes and transposons have been annotated.

Organelle Arabidopsis genomes have also been completely characterized (Unseld et al. 1997; Sato et al. 1999). The chloroplast genome contains 88 protein-coding and 45 RNA-coding genes, while the numbers for mitochondria are 122 and 24, respectively.

Summarizing, the last gene content prediction for the Arabidopsis genome is 27,029 protein-coding genes and 1,123 RNA-coding genes.

18.3.2 T-DNA as a Mutagen

Despite the availability of the whole genome sequence and the development of robust computational and technological tools for gene discovering, the most trustworthy way of determining the function of a gene is still the isolation and characterization of mutant alleles. Unfortunately, it is not easy to obtain mutant alleles in a directed way, since reports of knock-out mutants obtained by homologous recombination are extremely rare in plants (Kempin et al. 1997; Beetham et al. 1999). Although improvements to this procedure have been made in crops such as rice (Terada et al. 2002, 2007), the easiest way to find mutant alleles of a given gene is to look for them in nearly-saturation collections of sequenced-indexed mutants.

The T-DNA, which is naturally present in the Ti plasmid of the ground-living bacterium *Agrobacterium tumefaciens*, has been shown to be one of the most useful and powerful genetic tools for achieving transgenesis in plants (Gelvin 1998), as well as to find mutant alleles of a particular gene. This piece of DNA jumps from the bacterial Ti plasmid to the plant genome and stably integrates on it, making

almost any gene susceptible to being disrupted by a T-DNA insertion: the longer the gene, the most probability of a T-DNA landing on it in a transformation experiment. Assuming a median gene length of 2.1 kb, from the start to the stop codon of the genomic sequence, and random genomic T-DNA integration, Krysan and colleagues (1999) estimated that 280,000 T-DNA inserts were necessary to have a 99% chance of mutating a particular gene. A T-DNA insertion usually disrupts a gene generating recessive hypomorphic or null alleles. The extent of the damage is probably greater the further upstream in the gene the T-DNA is inserted. Moreover, it is more likely to disrupt a gene if the T-DNA integrates in an exon, since insertions in introns can be spliced out (Miesak and Coruzzi 2002). Integrations in gene promoters can also eliminate transcription.

Since the right and left border sequences of the T-DNA (RB and LB, respectively) are the only ones necessary to mediate the transfer event, the T-DNA can be manipulated to contain any desired sequence to be introduced in the Arabidopsis genome in order to facilitate transformant selection or to reveal or manipulate gene expression. The T-DNA used for insertional tagging usually includes a dominant selectable marker which confers resistance to antibiotics (typically kanamycin) or herbicides (generally gluphosinate). This allows the easy screening for transformants in mutagenesis experiments.

The main advantage of insertional screens is that the T-DNA-tagged genes can be isolated more easily than by using chemical or physical mutagens through the rapid identification of the genomic sequences flanking the T-DNA insertion by means of techniques such as plasmid rescue (Behringer and Medford 1992; Feldmann 1992; Okuley et al. 1994; Nakazawa et al. 2001), inverse PCR (IPCR; Gash et al. 1992; Ponce et al. 1998; Quesada et al. 1999) or TAIL (Thermal Asymmetric Interlaced) PCR (Liu and Whittier 1995; Liu et al. 1995; Sessions et al. 2002).

18.3.3 Collections of Insertional Mutants

The development of an easy protocol of *Agrobacterium*-mediated transformation, the floral dip method (Clough and Bent 1998), has permitted several Arabidopsis insertional mutant collections to be generated in the last decade. The goal of these collections is to reach a high degree of saturation of the plant genome with T-DNA insertions, and hence to obtain large numbers of indexed-mutants to perform high-throughput functional studies. These collections are generated in a similar way. A huge number, usually several thousands, of T_0 wild-type plants are transformed. T_1 seeds are collected and transformant T_1 plants, which are hemizygous for any integrated T-DNA, are selected and individually grown. Each T_1 line establishes a T_2 population in which resistant plants, homozygous or hemizygous for the T-DNA are selected, giving rise to several T_3 populations. The genomic insertion sites of T-DNA integration events, named as flanking sequence tags (FSTs), are determined in transformant individual lines and compiled in databases. This allows sequence homology searches (e.g. using

the BLAST program [http://www.ncbi.nlm.nih.gov/blast/Blast.cgi]) to be performed in order to find putative insertional alleles of the gene of interest.

There are three main collections of T-DNA induced mutants which can be used in Arabidopsis functional genomic research: the Salk, SAIL and GABI-Kat. All three of them were generated using the Columbia ecotype, probably the most frequently used by the Arabidopsis community and the one whose genome was sequenced in 2000. The SAIL (Syngenta Arabidopsis Insertion Library) collection of T-DNA indexed mutants was the first one to be generated (Sessions et al. 2002) and includes about 100,000 transformed lines, in which the integration sites have been determined in 52,964. The Salk collection, obtained at the SIGnAL (Salk Institute Genomic Arabidopsis Laboratory) in 2003 (Alonso et al. 2003), comprises 150,000 lines, including 225,000 estimated independent-T-DNA integration events. A total of 88,122 T-DNA insertion sites were mapped on the genome, revealing mutations in 21,799 AGI-predicted genes. The Salk collection today contains by far the highest number of insertional lines, it is the best characterized (with insertions mapped to date in about 150,000 lines) and the closest to saturation, for which reason it has become the most popular. The last of the big collections of T-DNA lines with mapped insertions is the GABI-Kat (Genome Analysis of the Plant Biological System), created at the Max Planck Institute of Germany (Rosso et al. 2003), which includes 59,979 transformant lines harboring insertions in 13,889 of the AGI-predicted genes. Other T-DNA mutant collections include the one generated at the Institute of Agronomics Research at Versailles (Samson et al. 2002) and that obtained by Rios and colleagues at the Max Planck Institute (Ríos et al. 2002).

The design of sophisticated T-DNA constructs has allowed development of other insertional mutagenesis approaches based on the use of transposable elements (transposon tagging). One of the most popular systems of transposon tagging in Arabidopsis takes advantage of the maize transposable element system *Activator*/*Dissociation* (*Ac/Ds*) (Sundaresan et al. 1995). In this approach, the *Ds* element usually contains a promoter or exon trap reporter gene, which makes it possible to determine whether the *Ds* integration happened near a *cis*-regulatory element (enhancer trap) or in a coding region (gene trap). The IMA (Parinov et al. 1999), RIKEN (Ito et al. 2002; Kuromori et al. 2004), CSHL (http://genetrap.cshl.org/), and EXOTIC (http://www.jic.ac.uk/science/cdb/exotic/index.htm) indexed-collections include around 65,000 lines whose *Ds* insertion sites have been mapped. Transposon tagging has been also achieved in Arabidopsis using the *Enhancer*/*Suppresor mutator* (*En*/*Spm*) transposable elements from maize (Tissier et al. 1999).

18.3.4 Obtaining and Characterizing T-DNA Tagged Mutants from Publicly Available Collections

In Arabidopsis research, the most usual way to carry out reverse genetics is to look for mutant alleles in publicly available tagged lines collections, in the expectation of finding a mutant phenotype. Although each of the above-mentioned high-throughput

mutagenesis projects has its own webpage, information about most of the insertional collections is gathered in several websites, such as the *Arabidopsis thaliana* Integrated Database (Pan et al. 2003; http://www.atidb.org/) or the SIGnAL on its T-DNA Express section (http://signal.salk.edu/cgi-bin/tdnaexpress). These sites allow all the information on the existing and available indexed-lines to be obtained in a friendly graphical way (Haag 2007). Thus, by introducing the AGI code of a gene whose function is going to be established, it is possible to obtain its relative genomic position, the insertional lines which are available and, for most of the lines, the orientation of the insertion and the genomic sequence adjacent to the LB of the T-DNA (Fig. 18.2a, b). To date, more than 379,000 mapped indexed-lines have been obtained (http://signal.salk.edu). With such a high density of mapped insertions it is quite probable to find at least one mutant allele in any given gene.

The next step in obtaining a mutant is to order the tagged lines from NASC (European Arabidopsis Stock Centre; http://arabidopsis.info/) or ABRC (Arabidopsis Biological Resource Center at the Ohio State University; http://www.biosci.ohio-state.edu/pcmb/Facilities/abrc/index.html) websites (Fig. 18.2c). These fundamental institutions receive seeds and DNA stocks donations from laboratories all over the world and maintain and distribute them to the Arabidopsis community. Most of the indexed-lines generated in the different mutagenesis projects are already available through NASC or ABRC. For a modest fee (a few pounds or dollars), it is possible to receive some T_3 seeds (Fig. 18.2d) which have to be screened to find the homozygous plants for the insertion, and checked whether they show a selectable mutant phenotype (Fig. 18.2e). Since resistance markers are frequently silenced in tagged lines, the most appropriate way to select the homozygous T_3 plants is to genotype the putative tagged locus by PCR using one gene-specific primer and one T-DNA specific-primer (Fig. 18.2f). The heritability of the phenotype of the T_3 mutant plants is checked in the T_4 generation (Fig. 18.2g). Taking into account the recessive nature of the insertional alleles (see Sect. 18.3.2) and the fact that only a few T_3 seeds per line are provided by NASC or ABRC, occasionally no T-DNA mutant homozygous plants will be found in T_3. In these cases, is required to generate T_4 populations from hemizygous T_3 individuals. The problem of identifying T-DNA homozygous mutant plants should be avoided thanks to the ambitious goal of the Salk collection: to obtain two homozygous insertional mutant lines for every Arabidopsis gene or, in other words, to achieve a collection of certified homozygous mutant lines. To date, around 18,000 of these lines, representing 13,000 genes, have already been deposited at the ABRC. Regardless of whether the line is certified homozygous or not, it is convenient to check by PCR and further sequencing the correct limits of the T-DNA integration.

Another consideration to take into account when working with T-DNA indexed mutants is the number of insertions per line. This, for the Salk collection, is 1.5 (Alonso et al. 2003). Since T-DNA tagged lines can bear more than one insertion, it is prudent to backcross the line at least once in order to eliminate unwanted T-DNA insertions (Fig. 18.2h). Moreover, T-DNA insertions are sometimes complex. They can display internal reorganizations or generate untagged mutations, such as deletions or chromosomal rearrangements due to abortive insertions (Tax and Vernon 2001). For this reason it is convenient to perform a co-segregation analysis of the mutant phenotype with a T-DNA insertion in the presumed tagged gene.

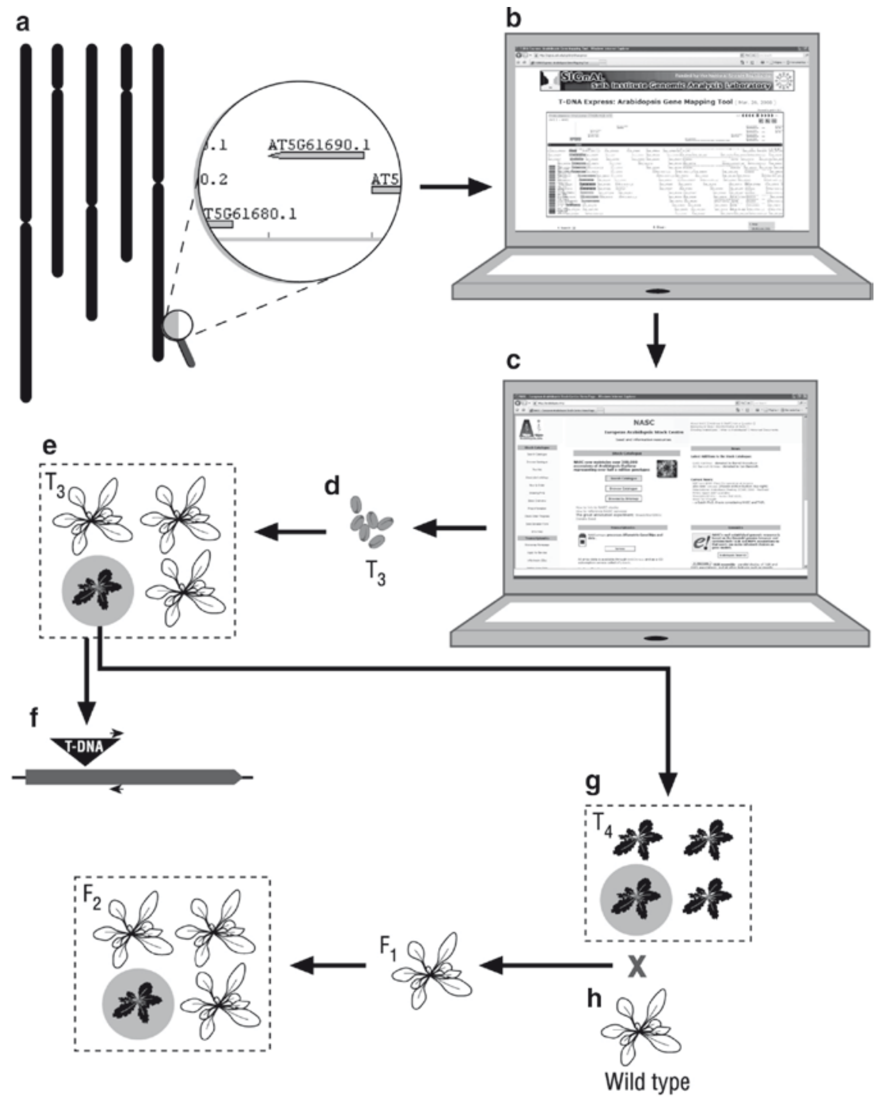


Fig. 18.2 Obtaining and characterizing tagged-mutant lines from publicly available insertional mutant collections. (a) In the reverse genetics approach, the functional characterization of any Arabidopsis gene starts from knowledge of its sequence. (b) The information about T-DNA insertions in the Arabidopsis genome, derived from different high-throughput tagging mutagenesis experiments, is accessible through several web pages (e.g. http://signal.salk.edu/cgi-bin/tdnaexpress). (c and d) Seeds of the publicly available insertional lines are provided by NASC in Europe and ABRC in the USA. (e) Assuming a recessive insertional mutation, wild-type (without T-DNA or hemizygous for the insertion) and mutant (homozygous for the insertion) plants can appear among the T_3 population. (f) T_3 plants are genotyped using combinations of T-DNA and gene specific primers (black arrows) in order to detect or confirm homozygous plants. The transcription unit of the gene disrupted by the insertion and the T-DNA are depicted as a gray arrowed box and an inverted triangle, respectively. (g) The heritability of the mutant phenotype is checked in the T_4 generation. (h) T_3 or T_4 mutant plants are backcrossed to eliminate T-DNA insertions not responsible for the mutant phenotype

The high number of T-DNA insertions spread along the Arabidopsis genome, obtained from the different high-throughput insertional mutagenesis experiments, makes it feasible to obtain a mutant allelic series for any given gene (see Sect. 18.2). The isolation of several mutant alleles of a particular gene giving rise to related phenotypes is usually clear evidence that the mutant phenotype is caused by the perturbation of that gene and that its function has been identified. Alternatively, if just one mutant allele leading to a mutant phenotype has been isolated, phenotypic rescue of the mutant with the wild-type allele using transgenesis is necessary in order to demonstrate gene function.

18.3.5 Activation Tagging Mutagenesis

To infer gene function when working with insertional loss-of-function mutants can be difficult in some cases. On the one hand, null alleles of some genes can be lethal in homozygosis if they code for essential housekeeping functions or are involved in embryo development. On the other hand, no mutant phenotype is observed in many loss-of-function mutants, because only around 35% of the predicted proteins of Arabidopsis are unique (The AGI 2000). In these cases, gain of function alleles can help to assign a function to a gene. For this, different T-DNA vectors, which usually include several copies of the enhancer element from the constitutively active promoter of the Cauliflower mosaic virus 35S gene (CaMV35S), have been developed (Weigel et al. 2000). Once these enhancers are integrated in the Arabidopsis genome they can cause transcriptional activation of close genes. Given that activated genes can be associated with a T-DNA insertion this approach has been named activation tagging. Integration of the activation construct can disrupt a gene but can also lead to enhanced or ectopic expression of the nearby gene, generating a new phenotype different from the loss-of-function one. Several collections of Arabidopsis activation-tagging lines have been generated (Weigel et al. 2000; Ichikawa et al. 2003) as a genome research resource. However, activation tagging can enhance the expression of more than one gene, hindering interpretation of the mutant phenotype (Ichikawa et al. 2003). The FOX (Full-length cDNA Over-eXpression) hunting system has recently been created to avoid this problem and a collection of Arabidopsis transgenic lines over-expressing full-length cDNAs has been obtained (Ichikawa et al. 2006).

18.3.6 Small RNAs as New Tools for Targeted Mutagenesis in Plants

The discovery in many eukaryotes of small molecules of RNA, known as silencing RNAs (sRNA), which mediate gene silencing at the transcriptional and post-transcriptional levels (Hannon et al. 2006), has opened up new revolutionary possibilities

of directed gene inactivation in plants and animals. In plants, different strategies have been devised to inactivate gene activity, taking advantage of this innate mechanism of gene silencing (Ossowski et al. 2008). These strategies usually include stable transformation with a transgene construct which produces or induces the formation of double stranded molecules of RNA (dsRNA). This gives rises to the generation of short interfering RNAs (siRNAs) which pair with complementary mRNA molecules, triggering a sequence-specific mRNA degradation that silences the targeted gene, a phenomenon termed RNA interference (RNAi) (Waterhouse and Helliwell 2003).

Compared to the insertional mutagenesis, RNAi-mediated gene silencing allows several sequence-related genes to be targeted simultaneously, making it possible to assess the phenotypic effect of interfering with the activity of all of them. However, RNAi can cause a range of effects in gene expression, ranging from complete inactivation to no reduction. The considerable variability in gene silencing limits high-throughput with RNAi because it requires several independent lines to be obtained and analysed, complicating interpretation of the results (Alonso and Ecker 2006). Additionally, it requires laborious plant transformation methods which are not available at the moment for many plant species.

MicroRNAs (miRNA) are a particular class of sRNAs which mediate the specific post-transcriptional control of endogenous mRNAs and originate from single-stranded precursor transcripts with imperfect fold-backs (Bartel 2004; Jover-Gil et al. 2005). In plants, miRNAs direct the cleavage of complementary mRNAs through RNA-induced silencing protein complexes (RISCs; Hammond et al. 2000). Recently, highly-specific artificial miRNAs (amiRNAs) have been developed in plants, using the sequence of endogenous miRNA precursors as a template (Parizotto et al. 2004), and they have been shown to effectively interfere with reporter or endogenous gene expression (Alvarez et al. 2006; Schwab et al. 2006). Once introduced by transgenesis in plants, amiRNAs could specifically target a particular gene or gene family, allowing functionally redundant genes or those with overlapping functions to be characterized (Alonso and Ecker 2006). Besides, the WDM (Web MicroRNA Designer) platform has been created to automate amiRNA design in Arabidopsis (Schwab et al. 2006) and in more than thirty plant species, including important crops such as rice or maize (Ossowski et al. 2008).

18.3.7 Tilling

Targeting Induced Local Lesion in Genomes (TILLING) is a highly sensitive method based on random chemical mutagenesis (by EMS), which allows the identification of point mutations in a genomic region of interest, after PCR screens of arrayed DNAs obtained from mutagenized individuals (McCallum et al. 2000a, b). Hence, this methodology combines traditional chemical mutagenesis with reverse genetics, providing a wide range of alleles very useful for dissecting gene function. A high-throughput TILLLING protocol for Arabidopsis has been generated (Till et al. 2006). Based on this, a service for the Arabidopsis research community,

known as the Arabidopsis TILLING Project (ATP) is available and users can request different EMS-derived alleles in genomic regions of interest (Till et al. 2003). TILLING has been applied to detect induced mutations in other plant species such as maize, rice, *Lotus japonicus* or soybean (see forward).

An exhaustive description of the TILLING method and its implementation in crops can be found in another chapter of this book.

18.4 Recent Progress in Crop Mutations and Functional Genomics

Despite the importance of Arabidopsis in plant biology, some desirable agronomic traits are not present in this model plant. Therefore, large-scale genomic projects of lesser-studied but economically important plant species would help to improve our knowledge about, for instance, how developmental and/or metabolic pathways are shared by different plants, which of them are species-specific and how they have been modified through evolution. The isolation of mutants in agronomic important plants is a major goal of classical crop mutation breeding programmes. Innovative breeding methods, together with new biotechnological tools and sequence genomic data, are a useful complement to classical breeding approaches.

One of the greatest advantages of mutagenesis as a method for dissecting gene function is that it can lead to different mutant alleles of a particular gene, causing different degrees of phenotypic trait modifications (see Sects. 18.2 and 18.3). To identify a gene whose mutations produce a phenotype different from the wild-type, the map-based cloning approach is the most effective strategy as long as the gene is untagged (see Sect. 18.2.1). However, positional cloning in agronomically important plants with large genomes, such as barley, maize or wheat is still challenging. Mutational approaches play an increasingly important role in creating crop varieties with modified traits. Yield, maturation, grain quality, biotic and abiotic stress resistance are the most manipulated/improved crop traits through mutagenesis. Mutation-derived varieties have been released for more than 175 crops and plant species, including cereals, grain legumes, oil and industrial crops, fruits, vegetable, horticultural and ornamental plants (Ahloowalia et al. 2004; Chopra 2005).

The potential of the different mutagenesis approaches combined with the genomic knowledge derived from plant genome sequencing projects is starting to make functional genomics possible in crops. According to the National Centre for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/genomes/static/gpstat.html), genome sequencing of five land plants, *Carica* papaya (370 Mb), *Physcomitrella patens* (510 Mb), *Populus trichocarpa* (480 Mb), *Ricinus communis* (400 Mb), and *Vitis vinifera* (500 Mb) is nearly completed. Sequencing projects of another 23 plant genomes are in progress, including agronomically interesting plants like *Brassica oleracea* (600 Mb), *Brassica rapa* (500 Mb), *Glycine max* (1,200 Mb), *Medicago trunculata* (500 Mb), *Solanum tuberosum* (840 Mb), *Sorghum bicolor* (760 Mb), and *Triticum aestivum* (16,000 Mb).

In the next sections, different methods and recently developed strategies for performing forward and reverse genetics in economically important crops are discussed.

18.4.1 Rice

Rice (*Oryza sativa*) serves as a model plant for cereal crop genomics. A high quality map-based draft of the nuclear genomic sequences of two divergent cultivars, *Oryza sativa* L. ssp. *japonica* and *Oryza sativa* L. ssp. *indica* has been obtained (Goff et al. 2002; Yu et al. 2002), providing a reference platform for genomics in cereals and grasses. Rice was chosen as a model because (a) its genome is the smallest of the graminaceous crops (389 Mb; International Rice Genome Sequencing Project 2005), (b) its global importance as a food crop and (c) the existence of a large-scale collaboration in the rice research community all around the world. The availability of a vast amount of information on rice germplasm resources, including mutants (Ryu et al. 2004; Sallaud et al. 2004), a large set of full-length cDNAs, more than 1,220,000 ESTs and DNA arrays (http://www.ncbi.nlm.nih.gov/dbEST; Lan et al. 2004) has helped in the design of novel rice varieties. Sequencing of the rice genome has led to the identification of more than 37,000 predicted genes (International Rice Genome Sequencing Project 2005), and one of the main goals of the rice community is to determine their function as quickly as possible.

In recent years, despite extensive efforts using chemical or irradiation mutagenesis, the function of only a handful of agriculturally important rice genes has been revealed, most of them through laborious map-based cloning (Jung et al. 2007). Some of these genes include *XANTHOMONAS ORYZAE PV.ORYZAE RESISTANCE 21* (*Xa21*), which is involved in bacterial resistance (Song et al. 1995), *MONOCULM 1* (*Moc1*) which controls tillering number (Li et al. 2003), *PISTILLATE FLORETS 9* (*Pi9*) and *2* (*Pi2*), which are required for fungal resistance (Qu et al. 2006; Zhou et al. 2006), *GIBBERELLIN-INSENSITIVE DWARF 1* (*Gid1*), *2* (*Gid2*) and *SLENDER RICE 1* (*Slr1*), which are involved in the gibberellin signaling pathway (Ikeda et al. 2001; Sasaki et al. 2003; Ueguchi-Tanaka et al. 2005), *SEMI-DWARF 1* (*Sd1*) required for gibberellin synthesis (Sasaki et al. 2002), *LOW SILICON RICE 1* (*Lsi1*), a silicon transporter (Ma et al. 2006), *QTL OF SEED SHATTERING IN CHROMOSOME 1* (*qSH1*) which controls grain abscission (Konishi et al. 2006) and *SUBMERGENCE TOLERANCE 1* (*Sub1*) (Xu et al. 2006). The knowledge of the function of these genes allows their introgression in relevant cultivars to produce varieties of agronomical importance. As an example of this, the *Sub1* gene has been introgressed into a widely grown Asian rice cultivar using marker-assisted selection. This new variety maintains the high yield and other agronomic properties of the recurrent parent, and is tolerant to submergence. Hence, it is expected to provide protection against damaging floods (Xu et al. 2006).

Insertional mutagenesis has been successfully applied in rice to obtain loss-of-function mutants and to identify the tagged genes (Jeon et al. 2000; Jung et al.

2003). High-throughput insertional mutagenesis in this plant has been carried out using (a) the endogenous retrotransposon *Tos17*, (b) the T-DNA of *A. tumefaciens*, (c) heterologous transposons such as *Ds* and *dSpm* and (d) the FOX system (see Sect. 18.3.5) (An et al. 2005; Kumar et al. 2005; Jeong et al. 2006; Hsing et al. 2007). An important experimental approach for rice reverse genetics is the screening of T-DNA insertional mutants through searches for FSTs in the databases (see Sect. 18.3.3). All the rice FSTs are publicly available at the Rice Functional Genomic Express Database (RiceGE; http://signal.salk.edu/cgi-bin/RiceGE) developed by The Salk Institute. Taken together, approximately 500,000 insertional mutant lines have been generated by the international rice research community (http://www.pi.csiro.au/fgrttpub/home.htm/, http://orygenesdb.cirad.fr/, http://tll.org.sg/sri.asp/, http://www.plb.ucdavis.edu/Labs/sundar/, http://www.niab.go.kr/).

Using chemical (such as diepoxybutane and EMS) and physical (such as fast neutrons and γ-rays) mutagens, around 60,000 mutants of Indica Rice I64 have been obtained. These lines are suitable for forward genetics and through TILLING for reverse genetics analysis (Wu et al. 2005). Finally, the RNAi technique has recently been used to dissect defense-signaling pathways in rice (Mei et al. 2007).

18.4.2 Maize

Maize (*Zea mays*) has been used as a model organism in Genetics since the nineteenth century (Coe et al. 2001) and mutants affecting the kernel (of great agronomic importance), meiosis, or genes essential for inflorescence development have been identified and characterized. Furthermore, maize has played a prominent role in studies of evolution and plant domestication (Candela and Hake 2008).

The completion of a draft sequence covering about 95% of the maize nuclear genome (2,400 Mb) was announced at the beginning of 2008 (http://www.maizesequence.org/index.html). The sequence of the maize genome containing 40,000–60,000 predicted genes represents an important resource for plant breeding and biotechnology companies, and will become a valuable reference for research.

Chemical mutagenesis in maize has been carried out by EMS. Researchers tend to prefer pollen grains rather than seeds for mutagenise because this guarantees that every mutation is a unique event (Neuffer 1982). As mentioned above, the main disadvantage when performing EMS mutagenesis in species with large genomes, such as maize, is the difficulty in identifying the mutated gene by map-based cloning. Nonetheless, the availability of the rice genomic sequence, which provides an anchor genome to potentially determine gene order in maize thanks to synteny conservation, and the draft sequence of the maize genome will accelerate map-based cloning (Hale et al. 2007). In the Maize Genetics and Genomics Database (MaizeGDB; http://www.maizegdb.org/ems-phenotype.php) a collection of maize phenotypic EMS-induced mutants is available for the research community.

Collections of insertional lines generated using the maize transposons *Ac/Ds* and *Spm* (Settles 2005) have been obtained. Nevertheless, the *Mutator* (*Mu*) family of

transposons is the most frequently used for gene tagging in maize because it offers several advantages, such as high mutation rates and the generation of many insertional copies per genome (Walbot et al. 2002), reducing the size of the population to be screened for the desired phenotypes (Lunde et al. 2003). Additionally, given that *Mu* elements are present in a high numbers in the maize genome, they are a useful tool for gene targeting. However, the segregation of numerous *Mu* elements complicates the isolation of simple knockout phenotypes (May et al. 2003). Moreover, transposable elements can generate unstable insertions, making it difficult to correlate a mutant phenotype with a single insertion site (May et al. 2003). To address the problem associated with a high copy number of transposons per genome and to facilitate the identification of insertions linked to the phenotype of interest, the *UniformMu* population, derived from an active *Mu* line introgressed into the W22 inbred line, has been developed (McCarty et al. 2005; Settles et al. 2007). Several collections of *Mu*-induced mutants and databases are available: the *RescueMu* generated with the *Mu1* element (http://www.maizegdb.org/rescuemu-phenotype.php), the *UniformMu* generated by the *UniformMu* Maize Project and the *P*hotosyntetic *M*utant *L*ibrary, a library of *Mu*-induced non-photosynthetic maize mutants (http://chloroplast.uoregon.edu/). In total, more than 150,000 mutagenized lines with stabilized insertions have been obtained in maize, through the use of transposon-efficient systems (Bensen et al. 1995; May et al. 2003; Fernandes et al. 2004; Stern et al. 2004; McCarty et al. 2005).

A maize TILLING project for reverse genetics in maize based on EMS-mutagenized populations in two genetic backgrounds, W22 and B73, has been already initiated (Till et al. 2004; Weil and Monde 2007).

Gene silencing using RNAi (see Sect. 18.3.6) has also been successfully used in maize to ascertain gene function, and more than 100 genes have already been targeted, providing a large number of stable transgenic lines (McGinnis et al. 2007).

18.4.3 Tomato

Tomato (*Lycopersicon esculentum*) is a model plant for fruit development, ripening and metabolism, for which extensive genomic data are available. With its small genome (950 Mb; Shibata 2005), numerous molecular markers, more than 257,000 available ESTs (http://sgn.cornell.edu/, http://www.tigr.org./tdb/lgi) and many mapped traits, tomato is one of the targets chosen in 2004 for genome sequencing (http://www.ncbi.nlm.nih.gov/genomes/static/gpstat.html). The International Solanaceae Genomics Project (SOL; http://www.sgn.cornell.edu/) was established with the goal of obtaining a high quality sequence of the tomato genome in order for use as a reference sequence for the Solanaceae. The function of most tomato genes (current rough estimations range from 30,000 to 35,000) is still unknown.

Mutational approaches have been highly exploited to study the genetic and molecular basis of many tomato traits (Menda et al. 2004) and mutants are still of direct use in classical breeding. In a recent mutagenesis experiment using EMS and

fast-neutrons, a total of 13,000 M_2 families were visually phenotyped (Menda et al. 2004), and more than 3,000 mutants from different experiments have been catalogued (http://zamir.sgn.cornell.edu/mutants). Most phenotypes among the "mutant library" are related to plant size, type of plant sterility, plant habit and leaf morphology. On the contrary, phenotypes related to fruit development, flower morphology and disease responses are less abundant (Menda et al. 2004). The rarest mutants (less than 1%) showed alterations in flowering time, flower color and inflorescence structure. Almost half of the described mutants displayed some rate of pleiotropy. More recently, 3,839 additional mutant plants have been obtained from EMS mutagenesis of the miniature dwarf tomato cultivar Micro-Tom (Watanabe et al. 2007), originally bred for home gardening purposes (Scott and Harbaugh 1989; Emannuel and Levy 2002). Moreover, successful application of EMS mutagenesis in tomato has yielded mutants with an increased resistance to the parasitic weed *Orobanche ramosa*. These results suggest that chemical mutagenesis can produce tomatoes with improved traits and can be useful in breeding programs focused on, for example, limiting the damage caused by parasitic plants (Kostov et al. 2007).

A physical mutagenesis using γ-rays have been also performed on Micro-Tom. A total of 6,347 lines were obtained and categorized according to morphological phenotypes. Segregation data suggest that most of the mutants have single recessive mutations (Matsukara et al. 2007). Because γ-ray mutagenesis causes mainly large-scale deletions (up to 6 Mb; Naito et al. 2005) (see Sect. 18.2), the reverse genetics approach is difficult to apply in the identification and isolation of the mutated genes.

Insertional mutagenesis, using the maize *Ac/Ds* transposable elements or the T-DNA, has been also applied in tomato. Several genes have been cloned by *Ac/Ds* transposon tagging. These include *Cf*-9 (Jones et al. 1994) and *Cf*-4 (Takken et al. 1998), both involved in resistance to *Cladosporium fluvum*, *DWARF* (*D*), encoding a cytochrome P450 homolog (Bishop et al. 1996), *DEFECTIVE CHLOROPLAST AND LEAVES* (*DCL*), controlling chloroplast development and palisade cell morphogenesis (Keddie et al. 1996), *FEEBLY* (*FB*) involved in metabolism and development (van der Biezen et al. 1996), and *DEFECTIVE EMBRYO AND MERISTEM* (*DEM*) required for correct organization of the shoot apical meristem in developing embryos and meristem maintenance in roots (Keddie et al. 1998).

By using activation-tagging (see Sect. 18.3.5), more than 10,000 tomato transgenic lines have been generated (Shibata 2005; Mathews et al. 2003) and around 13% of them show one or more visually observable phenotypes, demonstrating the effectiveness of this method for the induction of dominant mutations in tomato.

18.4.4 Other Important Crops and Species

Brassica species, such as *B. oleracea*, *B. rapa* and *B. napus*, are important vegetable and oilseed crops. As the second oilseed crop in the world, numerous studies have been performed on *B. napus* with the main focus on increasing yield (Zhang et al.

2006b). This research was carried out via genetic breeding, but with a significant contribution on the part of molecular biology and genetics. Currently, several databases for *Brassica* species are available: (a) one database functions on *B. napus* mitochondrial genome and the comparative analysis with Arabidopsis (Handa 2003), (b) the KEGG database providing EST and fatty acid metabolism resources (Kanehisa and Goto 2000), and (c) the databases incorporating Gene Ontology annotation, SSR molecular markers and expression profiles of thousands of uni-genes and MarkerQTL (Love et al. 2005; Erwin et al. 2007). Recently, the RAPESEED database (http://rapeseed.plantsignal.cn) was created to provide a platform for functional genomics studies of oilseed crops with emphasis on seed development and fatty acids. As regards functional genomics through mutagenesis, a total of 14,700 M_3 EMS-induced mutant plants were generated and a TILLING reverse genetics strategy was carried out to identify point mutations (Wu et al. 2008).

Barrel Medick, or simply Medicago (*Medicago truncatula*), is the reference for legume species. Its relatively small diploid genome (500 Mb), compared for instance with those of maize (2,400 Mb) or tomato (950 Mb) makes it useful for both genetics and genomics. Moreover, Medicago is suitable for reverse genetics approaches, since transformation is relatively easy to perform (Araújo et al. 2004; Crane et al. 2006). A number of useful genetic tools and genomic resources have been developed, such as high-density genetic and physical maps (Choi et al. 2004), different types of mutant populations (including EMS-, fast-neutrons-, and transposon-tagged mutants [Tadege et al. 2005]), as well as tools and protocols for transcriptome, proteome, and metabolome analysis (Gallardo et al. 2003; Watson et al. 2003; Barnett et al. 2004; Manthey et al. 2004; Broeckling et al. 2005). The most recent technique advance for Medicago functional genomics is the commercial availability of an Affymetrix GeneChip (http://www.affymetrix.com/products/arrays/specific/medicago.affx), which contains probe sets for the majority of the genes in this species.

The fact that grain traits can only be studied within cereals underlines the need for functional genomics resources for these species. Conventional cereal breeding is a time-consuming process, taking up to 12 years in some cases. Hence, breeders are interested in the use of new technologies that could make this process more rapid and effective. Genetic studies have resulted in the generation of multiple genetic maps with extensive number of molecular markers for many traits (Buerstmayer et al. 2002; Schmierer et al. 2005). Currently, the development of high-throughput methods for validating gene function in cereals is a priority. However, the lack of efficient transformation systems for the Triticeae and their large genomes are hindering the exploration of resources such as insertional mutagenesis. In contrast, chemical (using EMS and sodium azide) and physical mutagenesis (using fast neutrons and X- and γ-rays) provide a valuable alternative for obtaining mutant populations which can promote forward and reverse genetics in Triticeae.

A large collection of chemically and physically-induced barley (*Hordeum vulgare*) mutants has been collected lasting recent decades. Examples of viable

mutants are *eceriferum* (waxless) (Lundqvist and Lundqvist 1988), *breviaristatum* (short awn) (Kucera et al. 1975), *erectoides* (dense spike) (Hagberg 1958), *praematurum* (early-heading) (Hagberg 1958; Gustafsson et al. 1960) or *albostrians* (chlorophyll-deficient) (Hagemann and Scholz 1962). Pathogen-resistant (Jørgensen 1996; Zhang et al. 2006a) and anthocynin-deficient (Olsen et al. 1993) mutants have been also identified. Many of the lethal mutants isolated to date are affected in pigment biosynthesis, such as *albina*, *xantha*, *viridis* and *tigrina* (Hennigsen et al. 1993; Hanssonn et al. 1997).

The forward genetics approach has been employed in common wheat (*Triticum aestivum*) and durum wheat (*Triticum durum*) (Sakin et al. 2005). Some EMS-induced mutants have been reported, such as *waxy* and *low-amylose endosperm* (Yasui 2006), lesion mimic (Kamlofski et al. 2007) or several male-sterile mutants (Sasakuma et al. 1978). The thermotolerant mutants *tht* (Behl et al. 1986; Mullarkey and Jones 2000) have been obtained and analyzed thanks to the use of two chemical agents, EMS and sodium azide (Jones 1997) although the mutated genes have not yet been cloned. In wheat, a powerful reverse genetics approach was recently implemented through the combination of EMS mutagenesis and TILLING (see Sect. 18.3.7; Slade et al. 2005). Because of the tolerance of polyploid wheat to high mutation rates, this method is very efficient in identifying mutations in target genes. To date, wheat RNAi (see Sect. 18.3.6) has been successfully applied to targeting a wide range of genes, including those encoding transcription factors, storage proteins, starch biosynthetic enzymes and proteins involved in signaling and developmental processes (Loukoianov et al. 2005; Regina et al. 2006; Travella et al. 2006; Uauy et al. 2006; Yue et al. 2007).

Transposon elements, as major contributors and tools in genomic mutagenesis of many species, have also been used in cereals. Two-element tagging strategies, particularly the *Ac/Ds* maize system (see Sect. 18.3.5) have proven effective in barley (McElroy et al. 1997) and wheat (Takumi et al. 1999).

Many research resources have been established for barley, wheat, rye, oat and sugarcane, such as EST projects, the development of markers and their implementation in marker-assisted breeding, genomics, and comparative mapping (http://barleygenomics.wsu.edu/, http://wheat.pw.usda.gov/ggpages/genomics.shtml).

18.4.5 Weeds

Despite the fact that weeds cause large yield losses in crops, there is little known about the genomics of economically relevant weed species. Although the whole-genome sequencing of the model "weed" Arabidopsis is largely inapplicable to any other weed species, the weed research community is taking full advantage of the Arabidopsis genome knowledge and trying to employ some of the information to weed genomics. Apart from Arabidopsis, other reference genomes like wheat, sorghum and rice are used to study different biological processes in weeds (Foley 2002; Chao et al. 2005). Comparative genomics, mutagenesis strategies, DNA

microarrays, ESTs databases and transformation systems provide the most practical tools for the characterization of interesting genes underlying specific weed-like characteristics (Foley 2002; Fukao et al. 2003; Borsics et al. 2002).

Forward and reverse genetics approaches have been used in weeds (Gu et al. 2004; Horvath and Anderson, 2002; Horvath et al. 2003), although there are some limitations to their application due to the lack of high-throughput transformation technologies, which limits the production of large mutant libraries. Reverse genetics tools such as TILLING (and especially EcoTILLING), and RNAi could be potentially performed in weed genomics (Waterhouse and Heliwell, 2003; Comai et al. 2004).

The selection of weed candidates for further genomic studies is crucial. Currently, four model systems are studied: leafy spurge (*Euphorbia esula*), weedy rice (*Oryza sativa* f. *spontanea*), Johnsongrass (*Sorghum halepense*) and thale cress (*Arabidopsis thaliana*). Another candidate recommended for the study of weed biology through genomics is the consummate weed waterhemp (*Amaranthus* spp.) because of its rapid evolution, economic importance, and relatively small genome (657 Mb) (Basu et al. 2004; Chao et al. 2005). It has been reported that this plant species has developed resistance to several herbicides (Heap 2004). The resistance of weeds to herbicides is a characteristic that can potentially cause significant economic losses. Three hundred and seventeen unique herbicide resistant biotypes, including 183 weed species, have been reported to date (http://www.weedresearch.com/). Recently, a γ-ray mutagenesis approach has been used to generate mutant plants of *Amaranthus tricolor* displaying increased drought tolerance and productivity (Slabbert 2007). Other γ-rays mutants have also been isolated in the related species *Amaranthus cruentus* showing enhanced grain yield (Gajdošová et al. 2007).

18.5 Conclusions and Perspectives

The dawning of the genomic era at the beginning of the twenty-first century has produced a revolution in Biology comparable to those generated by the discovery of the molecular nature of genetic material, the tri-dimensional structure of DNA or the development of molecular tools and technologies for performing genetic engineering. Today, the huge amount of data derived from the analysis of completely sequenced plant genomes, such as Arabidopsis and rice, and from those well on the way to being sequenced, are providing the research community with invaluable resources for carrying out functional and comparative genomic studies. Functional genomics and systems biology (Kirschner 2005; Yuan et al. 2008) are facilitating the identification of gene networks that control the genetic variation of agronomically important traits.

In this chapter, we have attempted to highlight the important role played by mutagenesis in the dissection of gene function through the identification and characterization of mutants. The powerful methodologies and technologies depicted in this chapter, such as chemical, physical or insertional mutagenesis, activation-tagging, TILLING or amiRNAs are incessantly contributing to our understanding of how genes function in

different biological processes. Many of these methodologies are the basis of forward and reverse genetics approaches, which, first developed in Arabidopsis, are now being applied to agronomically important plants. The generation and further analysis of mutant lines in crops and comparison of the effects of the perturbation of homologous genes in different plant species is already providing information about the conservation or modification of gene function during evolution.

World food demand is constanly increasing mainly due to human overpopulation. Moreover, rapid climate change is forcing the development of crop varieties able to deal with this enviromental challenge. Gene mutations leading to trait improvement in existing cultivars or the breeding of new varieties obtained from mutagenesis has contributed to enhance germplasm technology. In many cases, gene mutation has triggered improvements in the value of a cultivar or a new crop. With the information regarding the location and function of gene(s) encoding for valuable traits, scientists will be well equipped to undertake the efficient implementation of genomic-assisted crop improvements and the breeding of new varieties with precise combinations of desirable traits.

Acknowledgements The authors wish to thank Pablo García-García for the graphic design of the figures.

References

Adamczyk BJ, Lehti-Shiu MD, Fernandez DE (2007) The MADS domain factors AGL15 and AGL18 act redundantly as repressors of the floral transition in Arabidopsis. Plant J 50:1007–1019

Aharoni A, Ric de VIs CH, Verhoeven HA et al (2002 Nontargeted metabolome analysis by use of Fourier transform ion cyclotron mass spectrometry. OMICS 6:217–234

Ahloowalia BS, Maluszynski M, Nichterlein K (2004) Global impact of mutation-derived varietes. Euphytica 135:187–204

Alba R, Feil Z, Payton P et al (2004) ESTs, cDNA microarrays, and gene expression profiling: tools for dissecting plant physiology and development. Plant J 39:697–714

Allen RS, Li J, Stahle MI, Dubroue` A et al (2007) Genetic analysis reveals functional redundancy and the major target genes of the Arabidopsis miR159 family. Proc Natl Acad Sci USA 104:16371–16376

Alonso JM, Ecker JR (2006) Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in Arabidopsis. Nat Rev Genet 7:524–536

Alonso JM, Stepanova AN, Leisse TJ et al (2003) Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science 301:653–657

Alvarez JP, Pekker I, Goldshmidt A et al (2006) Endogenous and synthetic microRNAs stimulate simultaneous, efficient, and localized regulation of multiple targets in diverse species. Plant Cell 18:1134–1151

An G, Lee S, Kim SH et al (2005) Molecular genetics using T-DNA in rice. Plant Cell Physiol 46:14–22

Araújo S, Duque ASRLA, dos Santos DMMF et al (2004) An efficient transformation method to regenerate a high number of transgenic plants using new embryogenic line of *Medicago truncatula* cv. Plant Cell Tissue Organ Cult 78:123–131

Barnett MJ, Toman CJ, Fisher RF et al (2004) A dual-genome Symbiosis Chip for coordinate study of signal exchange and development in a prokaryote-host interaction. Proc Natl Acad Sci USA 101:16636–16641

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297

Basu C, Halfhill MD, Mueller TC et al (2004) Weed genomics: new tools to understand weed biology. Trends Plant Sci 9:391–398

Beetham PR, Kipp PB, Sawycky XL et al (1999) A tool for functional plant genomics: chimeric RNA/DNA oligonucleotides cause in vivo gene-specific mutations. Proc Natl Acad Sci USA 96:8774–8778

Behl RK, Sheoran IS, Kuhad MS et al (1986) Performance of radiation-induced high temperature tolerant spring wheat mutant under different environmental conditions. Proceedings of international seminar on water management in arid and semi-arid Zones, UNESCO, pp 260–273

Behringer FJ, Medford JI (1992) A plasmid rescue technique for the recovery of plant DNA disrupted by T-DNA insertion. Plant Mol Biol Reptr 10:190–198

Bender W, Spierer P, Hogness DS (1983) Chromosomal walking and jumping to isolate DNA from the Ace and rosy loci and the bithorax complex in Drosophila melanogaster. J Mol Biol 168: 17–33

Benfey PN, Linstead PJ, Roberts K et al (1993) Root development in Arabidopsis: four mutants with dramatically altered root morphogenesis. Development 119:57–70

Bensen RJ, Johal GS, Crane VC et al (1995) Cloning and characterization of the maize *An1* gene. Plant Cell 7:75–84

Berná G, Robles P, Micol JL (1999) A mutational analysis of leaf morphogenesis in Arabidopsis. Genetics 152:729–742

Bishop GJ, Harrison K, Jones JDG (1996) The tomato *Dwarf* gene isolated by heterologous transposon tagging encodes the first member of a new cytochrome P450 family. Plant Cell 8:959–969

Borsics T, Mihálka V, Oreifig AS et al (2002) Methods for genetic transformation of the parasitic weed dodder (Cuscuta trifolii bab. Et Gibs) and fro PCR-based detection of early transformation events. Plant Sci 162:193–199

Bowman JL, Smyth DR, Meyerowithz EM (1989) Genes directing flower development in Arabidopsis. Plant Cell 1:37–52

Broeckling CD, Huhman DV, Farag MA et al (2005) Metabolic profiling of *Medicago truncatula* cell cultures reveals the effects of biotic and abiotic elicitors on metabolism. J Exp Bot 56:323–336

Brown PO, Botstein D (1999) Exploring the new world of the genome with DNA microarrays. Nat Genet 21:33–37

Bruggemann E, Handwerger K, Essex C et al (1996) Analysis of fast neutron-generated mutants at the Arabidopsis thaliana HY4 locus. Plant J 10:755–760

Brumbarova T, Matros A, Mock HP et al (2008) A proteomic study showing differential regulation of stress, redox regulation and peroxidase proteins by iron supply and the transcription factor FER. Plant J 54:321–334

Buerstmayer H, Lemmens M, Hartl L et al (2002) Molecular mapping of QTLs for Fusarium head blight resistance in spring wheat. I. Resistance to fungal spread (Type II resistance). Theor Appl Genet 104:84–91

Candela H, Hake S (2008) The art and design of genetic screens: maize. Nat Rev Genet 9:192–203

Cecchini E, Mulligan BJ, Covey SN et al (1998) Characterization of gamma irradiation-induced deletion mutations at a selectable locus in *Arabidopsis.* Mutat Res 401:199–206

Chandler J, Wilson A, Dean C (1996) Arabidopsis mutants showing an altered response to vernalization. Plant J 10:637–644

Chao WS, Horvath DP, Anderson JV et al (2005) Potential model weeds to study genomics, ecology, and physiology in the 21st century. Weed Sci 53:929–937

Choi HK, Ki DJ, Uhm TS et al (2004) A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with M. sativa. Genetics 166:1463–1502

Chopra VL (2005) Mutagenesis: investigating the process and processing the outcome for crop improvement. Curr Sci 89:353–359

Chory J, Ecker JR, Briggs S et al (2000) National Science Foundation-Sponsored Workshop Report: "The 2010 Project" functional genomics and the virtual plant. A blueprint for understanding how plants are built and how to improve them. Plant Physiol 123:423–426

Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J 16:735–743

Coe EHJ (2001) The origins of maize genetics. Nat Rev Genet 2:898–905

Comai L, Young K, Till BJ et al (2004) Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. Plant J 37:778–786

Crane C, Wright E, Dixon RA et al (2006) Transgenic *Medicago trunculata* plantsobtained from *Agrobacterium tumefaciens*-transformed roots and *Agrobacterium rhizogenes*-transformed hairy roots. Planta 223:1344–1354

Ditta G, Pinyopich A, Robles P et al (2004) The *SEP4* Gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. Curr Biol 14:1935–1940

Emannuel E, Levy AA (2002) Tomato mutants as tools for functional genomics. Curr Biol 5:112–117

Erwin TA, Jewell EG, Love CG et al (2007) BASC: an integrated bioinformatics system for *Brassica* research. Nucleic Acids Res 35:870–873

Eshed Y, Baum SF, Perea JV et al (2001) Establishment of polarity in lateral organs of plants. Curr Biol 11:1251–1260

Feldmann KA (1992) T-DNA insertion mutagenesis in Arabidopsis, seed infection/transformation. In: Koncz C, Chua NH, Schell J (eds) Methods in Arabidopsis research. World Scientific, London, pp 16–82

Fernandes J, Dong Q, Schneider B, et al. (2004) Genome-wide mutagenesis of *Zea mays* L. using *Rescue*Mu transposons. Genome Biol 5:R82

Foley ME (2002) Weeds, seeds and buds: opportunities and systems for dormancy investigations. Weed Sci 50:267–272

Frick UB, Schaller A (2002) cDNA microarray analysis of fusicoccin-induced changes in gene expression in tomato plants. Planta 216:83–94

Fukao T, Ida S, Rumpho ME et al (2003) Differential gene expression of the α-chain of mitochondrial H^+ - transporting ATP synthase between dormant and non-dormant seeds of paddy *Echinochloa* weeds. Weed Biol Manage 3:15–20

Gajdošová A, Libiaková G, Fejér J (2007) Improvement of selected Amaranthus cultivars by means of mutation induction and biotechnological approaches. In: Ochatt S, Jain SM (eds) Breeding of neglected and under-utilized crops, spices and herbs. Science Publisher, Enfield (NH), Jersey, Plymouth 151–169

Gallardo K, Signor Cl, Vandekerckhove J et al (2003) Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. Plant Physiol 133:664–682

Gash A, Aoyama T, Foster R et al (1992) Gene isolation with the polymerase chain reaction. In: Koncz C, Chua NH, Schell J (eds) Methods in Arabidopsis research. World Scientific, London, pp 16–82

Gelvin SB (1998) The introduction and expression of transgenes in plants. Curr Opin Biotechnol 9:227–232

Goff SA, Ricke D, Lan THP et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92–100

Gong JM, Waner DA, Horie T, et al. (2004) Microarray-based rapid cloning of an ion accumulation deletion mutant in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 43:15404–15409

Grinikh LI, Shevchenko VV (1976) Number of initial cells forming the reproductive tissue in vegetating Arabidopsis thaliana plants. Genetika, Moscow, Russia 12:690–692

Gu XY, Kianian SF, Foiley ME (2004) Multiple loci and epistases control genetic variation for seed dormancy in weedy rice (Oryza sativa). Genetics 166:1503–1516

Gustafsson Å, Hagberg A, Lundqvist U (1960) The induction of early mutants in Bonus barley. Hereditas 46:675–699

Haag ES (2007) Dial-a-mutant: web-based knockout collections for model organisms. Biol Cell 99:343–347

Haas BJ, Wortman JR, Ronning CM, et al. (2005) Complete reannotation of the Arabidopsis genome: methods, tools, protocols and the final release. BMC Biol 3:7

Hagberg A (1958) Cytogenetik einger Gerstenmutanten. Züchter 28:32–36

Hagemann R, Scholz F (1962) Ein fall geninduzierter Mutationen des Plasmotypus bei gerste. Der Zücher 32:50–59

Hale CJ, Stonaker JL, Gross SM, et al. (2007) A novel Snf2 protein maintains *trans*-generational regulatory states established by paramutation in maize. PLoS Biol 5:e275

Hammond SM, Bernstein E, Beach D, et al. (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature 404:293–296

Handa H (2003) The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. Nucleic Acid Res 31:5907–5916

Hannon GJ, Rivas FV, Murchison EP, et al. (2006) The expanding universe of noncoding RNAs. Cold Spring Harb Symp Quant Biol 71:551–564

Hanssonn M, Gough SP, Kannagara CG, et al. (1997) Analysis of RNA and enzymes of potential importance for regulation of 5-aminolevulinic acid synthesis in the protochlorophyllide accumulating barley mutant tigrina-d12. Plant Physiol Biochem 35:827–836

Heap I (2004) The international survey of herbicide resistant weeds. www.weedscience.com

Hennigsen KW, Boyton JE, von Wettstein D (1993) Mutants at Xantha and Albino loci in relation to chloroplast biogenesis in barley (Hordeum vulgare L.). The Royal Danish Academy of Sciences and Letters, Copenhagen

Horvath DP, Anderson JV (2002) A molecular approach to understanding root bud dormancy in leafy spurge. Weed Sci 50:227–231

Horvath DP, Schaffer R, Wisman E (2003) Identification of genes induced in emerging tillers of wild oat (*Avena sativa*) using *Arabidopsis* microarrays. Weed Sci 51:503–508

Hsing Yi, Chern ChG, Fan MJ, et al. (2007) A rice gene activation/knockout mutant resource for high throughput functional genomics. Plant Mol Biol 63:351–364

Hülskamp M, Misŕa S, Jürgens G (1994) Genetic dissection of trichome cell development in Arabidopsis. Cell 76:555–566

Ichikawa T, Nakazawa M, Kawashima M, et al. (2003) Sequence database of 1172 T-DNA insertion sites in Arabidopsis activation-tagging lines that showed phenotypes in T1 generation. Plant J 36:421–429

Ichikawa T, Nakazawa M, Kawashima M, et al. (2006) The FOX hunting system: an alternative gain-of-function gene hunting technique. Plant J 48:974–985

Ikeda A, Ueguchi-Tanaka M, Sonoda Y, et al. (2001) *S*lender Rice, a constitutive gibberellin response mutant, is caused by a null mutation of the SLR1 gene, an ortholog of the height-regulating gene GAI/RGA/RHT/D8. Plant Cell 13:999–1010

International Rice Genome Sequencing Project (2005) The map based sequence of the rice genome. Nature 436:793–800

Ito T, Motohashi R, Kuromori T, et al. (2002) A new resource of locally transposed dissociation elements for screening gene-knockout lines in silico on the Arabidopsis genome. Plant Physiol 129:1695–1699

Jander G, Norris SR, Rounsley SD, et al. (2002) *Arabidopsis* map-based cloning in the post-genome era. Plant Physiol 129:440–450

Jeon JS, Lee S, Jung KH, et al. (2000) T-DNA insertional mutagenesis for functional genomics in rice. Plant J 22:561–570

Jeong DH, An S, Kang HG, Moon S, et al. (2002) T-DNA insertional mutagenesis for activation tagging in rice. Plant Physiol 130:1636–1644

Jeong DH, An S, Park S, et al. (2006) Generation of a flanking sequence-tag database for activation-tagging lines in japonica rice. Plant J 45:123–132

Jones DA, Thomas CM, Hammond-Kosack KE, et al. (1994) Isolation of tomato *Cf*-9 gene for resiatance to *Cladosporium fulvum* by transposon tagging. Science 266:789–792

Jones P (1997) Induced mutagenesis of wheat. In: Koebner R (ed) Manual for cytogenetic techniques, European Union, pp 2–7

Jørgensen JH (1996) Effect of three suppressors on the expression of powdery mildew resistance genes in barley. Genome 39:492–498

Jover-Gil S, Candela H, Ponce MR (2005) Plant microRNAs and development. Int J Dev Biol 49:733–744
Jung KH, An G, Ronald PC (2007) Towards a better bowl of rice: assigning function to tens of thousands of rice genes. Nat Rev Genet 9:91–101
Jung KH, Hur J, Ryu CH et al (2003) Characterization of a rice chlorophyll-deficient mutant using the T-DNA gene-trap system. Plant Cell Physiol 44:463–472
Jürgens G (1994) Pattern formation in the embryo. In: Meyerowitz EM, Somerville CR (eds) Arabidopsis. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 297–312
Jürgens G, Mayer U, Torres Ruiz RA et al (1991) Genetic analysis of pattern formation in the *Arabidopsis.* Dev Suppl 1:27–38
Kamlofski CA, Antonelli E, Bender C et al (2007) A lesion-mimic mutant of wheat with enhanced resistance to leaf rust. Plant Pathol 5:46–54
Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28:27–30
Keddie JC, Caroll BJ, Thomas CM et al (1998) Transposon tagging of the *Defective embryo and meristem* gene of tomato. Plant Cell 10:877–888
Keddie JS, Carroll B, Jones JDG et al (1996) The *DCL* gene of tomato is required for chloroplast development and palisade cell morphogenesis in leaves. EMBO J 15:4208–4217
Kempin SA, Liljegren SJ, Block LM et al (1997) Targeted disruption in Arabidopsis. Nature 389:802–803
Kim YS, Schumaker KS, Zhu JK (2006) EMS mutagenesis of Arabidopsis. Methods Mol Biol 323:105–113
Kirschner MW (2005) The meaning of systems biology. Cell 121:503–504
Konishi S, Izawa T, Lin SY et al (2006) An SN caused loss of seed shattering during rice domestication. Science 312:1392–1396
Koorneef M, Dellaert LWM, van der Veen JH (1982) EMS- and radiation-induced mutation frequencies at individual loci in *Arabidopsis thaliana* (L.) Heynh Mutat Res 93:109–123
Kostov K, Batchvarova R, Slavov S (2007) Application of Chenical mutagenesis to increase the resistance of tomato to *Orobanche ramosa* L. Bulgarian J Agr Sci 13:505–513
Krysan PJ, Young JC, Sussman MR (1999) T-DNA as an insertional mutagen in Arabidopsis. Plant Cell 11:2283–2290
Kucera L, Lundqvist U, Gustafsson A (1975) Induction of breviaristatum mutants in barley. Hereditas 80:263–278
Kumar CS, Wing RA, Sundaresan V (2005) Efficient insertional mutagenesis in rice using the maize En/Spm elements. Plant J 44:879–892
Kumaran MK, Bowman JL, Sundaresan V (2002) *YABBY* polarity genes mediate the repression of *KNOX* homeobox genes in Arabidopsis. Plant Cell 14:2761–2770
Kuromori T, Hirayama T, Kiyosue Y et al (2004) A collection of 11,800 single-copy Ds transposon insertion lines in Arabidopsis. Plant J 37:897–905
Lan L, Chen W, Lai Y et al (2004) Monitoring of gene expression profiles and isolation of candidate genes involved in pollination and fertilization in rice (*Oryza sativa* L.) with a 10 K cDNA microarray. Plant Mol Biol 54:471–487
Léon-Kloosterziel KM, van de Bunt GA, Zeevaart JA et al (1996) Arabidopsis mutants with a reduced seed dormancy. Plant Physiol 110:233–240
Li SL, Rédei GP (1969) Estimation of mutation rates in autogamous diploids. Radiat Bot 9:125–131
Li X, Zhang Y (2002) Reverse genetics by fast neutron mutagenesis in higher plants. Funct Integr Genomics 2:254–258
Li X, Qian Q, Zhiming F et al (2003) Control of tillering in rice. Nature 422:618–621
Li X, Song Y, Century K et al (2001) Fast neutron deletion mutagenesis-based reverse genetics system for plants. Plant J 27:235–242
Liu YG, Whittier RF (1995) Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics 25:674–681

Liu YG, Mitsukawa N, Oosumi T et al (1995) Efficient isolation and mapping of Arabidopsis thaliana T-DNA insert junctions by thermal asymmetric interlaced PCR. Plant J 8:457–463
Loukoianov A, Yan L Blechl A et al (2005) Regulation of *VRN*-1 vernalization genes in normal and transgenic polyploidy wheat. Plant Physiol 138:2346–2373
Love CG, Robinson AJ, Lim GAC et al (2005) *Brassica* ASTRA: an integrated database for *Brassica* genomic research Nucleic Acids Res 33:656–659
Lunde CF, Morrow DJ, Roy LM et al (2003) Progress in maize gene discovery: a project update. Funct Integr Genomics 3:25–32
Lundqvist U, Lundqvist A (1988) Induced *intermedium* mutants in barley: origin, morphology, and inheritance. Hereditas 108:13–26
Ma JF, Yamaji N, Mitani M et al (2006) A silicon transporter in rice. Nature 440:688–691
Manthey K, Krajinski F, Hohnjec N et al (2004) Transcriptome profiling in root nodules and arbuscular mycorrhiza identifies a collection of novel genes induced during *Medicago truncatula* root endosymbioses. Mol Plant-Microbe Interact 17:1063–1077
Mathews H, Clendennen SK, Caldwell et al (2003) Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. Plant Cell 15:1689–1703
Matsukara Ch, Yamaguchi I, Inamura M et al (2007) Generation of gamma irradiation-induced mutant lines of the miniature tomato (*Solanum lycopersicum* L.) cultivar "Micro-tom". Plant Biotech 24:39–44
May BP, Liu H, Vollbrecht E et al (2003) Maize-targeted mutagenesis: a knockout resource for maize. Proc Natl Acad Sci USA 100:11541–11546
Mayer U, Torres-Ruíz RA, Berleth T et al (1991) Mutations affecting body organization in the Arabidopsis embryo. Nature 353:402–407
McCallum CM, Comai L, Greene EA et al (2000a) Targeted screening for induced mutations. Nat Biotechnol 18:455–457
McCallum CM, Comai L, Greene EA et al (2000b) Targeting induced local lesions in genomes (TILLING) for plant functional genomics. Plant Physiol 123:439–442
McCarty D, Settles AM, Suzuki M et al (2005) Steady-state transposon mutagenesis in inbred maize. Plant J 44:52–61
McElroy DJ, Louwerse SM, McElroy SM et al (1997) Development of a simple transient assay for *Ac-Ds* activity in cells of intact barley tissue. Plant J 11:157–165
McElver J, Tzafrir I, Aux G et al (2001) Insertional mutagenesis of genes required for seed development in Arabidopsis thaliana. Genetics 159:1751–1763
McGinnis K, Murphy N, Carlson AR et al (2007) Assessing the efficiency of RNA interference for maize functional genomics. Plant Physiol 143:1441–1451
McKendry P (2002) Energy production from biomass (part 2): overview of biomass. Bioresource Technol 83:47–56
Mei CH, Zhou X, Yang Y (2007) Use of RNA interference to dissect defense-signaling pathways in rice. Plant-Pathogen Interact 354:161–171
Menda N, Semel Y, Peled D et al (2004) In silico screening of a saturated mutation library in tomato. Plant J 38:861–872
Meyerowitz EM (1994) Pattern formation in plant development: four vignettes. Curr Opin Genet Dev 4:602–608
Miesak BH, Coruzzi GM (2002) Molecular and physiological analysis of Arabidopsis mutants defective in cytosolic or chloroplastic aspartate aminotransferase. Plant Physiol 129:650–660
Mullarkey M, Jones P (2000) Isolation and analysis of thermotolerant mutants of wheat. J Exp Botany 51:139–146
Müller AJ (1965) The chimerical structure of M1 plants and its bearing on the determination of mutation frequencies in Arabidopsis. In: Veleminsky J, Gichner T (eds) Induction of mutations and the mutation process. Publishing House of the Czech Academy of Sciences, Prage, Czchoslovakia, pp 46–52
Naito K, Kusaba M, Shikazono N et al (2005) Transmissible and nontransmissible mutations induced by irradiating *Arabidopsis thaliana* pollen with γ-rays and carbon ions. Genetics 169:881–889

Nakazawa M, Yabe N, Ichikawa T et al (2001) DFL1, an auxin-responsive GH3 gene homologue, negatively regulates shoot cell elongation and lateral root formation, and positively regulates the light response of hypocotyl length. Plant J 25:213–221

Neuffer MG (1982) Growing maize for genetic purposes. In: Sheridan WF (ed) Maize for biological research. Univ North Dakota Press, Grand Forks, 19–30

Niyogi KK, Grossman AR, Björkman O (1998) Arabidopsis mutants define a central role for the Xanthophyll cycle in the regulation of photosynthetic energy conversión. Plant Cell 10:1121–1134

Ochando I, Jover-Gil S, Ripoll JJ et al (2006) Mutations in the microRNA complementarity site of the *INCURVATA4* gene perturb meristem function and adaxialize lateral organs in Arabidopsis. Plant Physiol 141:607–619

Okuley J, Lightner J, Feldmann K et al (1994) Arabidopsis *FAD2* gene encodes the enzyme that is essential for polyunsaturated lipid síntesis. Plant Cell 6:147–158

Olsen O, Wang X, von Wettstein D (1993) Sodium azide mutagenesis: preferential generation of A.T→G.C transitions in the barley Ant18 gene. Proc Natl Acad Sci USA 90:8043–8047

Ossowski S, Schwab R, Weigel D (2008) Gene silencing in plants using artificial microRNAs and other small RNAs. Plant J 53:674–690

Østergaard L, Yanofsky MF (2004) Establishing gene function by mutagenesis in Arabidopsis thaliana. Plant J 39:682–696

Pan X, Liu H, Clarke J et al (2003) ATIDB: Arabidopsis thaliana insertion database. Nucleic Acids Res 31:1245–1251

Parinov S, Sevugan M, Ye D et al (1999) Analysis of flanking sequences from dissociation insertion lines: a database for reverse genetics in Arabidopsis. Plant Cell 11:2263–2270

Parizotto EA, Dunoyer P, Rahm N et al (2004) In vivo investigation of the transcription, processing, endonucleolytic activity, and functional relevance of the spatial distribution of a plant miRNA. Genes Dev 18:2237–2242

Peters JL, Cnops G, Neyt P et al (2004) An AFLP-based genome-wide mapping strategy. Theor Appl Genet 108:321–327

Ponce MR, Quesada V, Micol JL (1998) Rapid discrimination of sequences flanking and within T-DNA insertions in the *Arabidopsis* genome. Plant J 14:497–501

Ponce MR, Robles P, Micol JL (1999) High-throughput genetic mapping in Arabidopsis thaliana. Mol Gen Genet 261:408–415

Ponce MR, Robles P, Lozano FM et al (2006) Low-resolution mapping of untagged mutations. Methods Mol Biol 323:105–113

Qu S, Desai A, Wing R et al (2008) A versatile transposon-based activation tag vector system for functional genomics in cereals and other monocot plants. Plant Physiol 146:189–199

Qu S, Liu G, Zhou B et al (2006) The broad-spectrum blast resistance gene *Pi9* encodes a nucleotide-binding site-leucine-rich repeat protein and is a member of a multigene family in rice. Genetics 172:1901–1914

Quesada V, Ponce MR, Micol JL (1999) OTC and AUL1, two convergent and overlapping genes in the nucleargenome of Arabidopsis thaliana. FEBS Lett 461:101–106

Quesada V, Ponce MR, Micol JL (2000) Genetic analysis of salt tolerant mutants in *Arabidopsis thaliana*. Genetics 154:421–436

Rédei GP, Koncz C (1992) Classical mutagenesis. In: Koncz C, Chua NH, Schell J (eds) Methods in Arabidopsis research. World Scientific, London, pp 16–82

Regina A, Bird A, Topping D et al (2006) High-amylose wheat generated by RNA interference improves indices of large-bowel health in rats. Proc Natl Acad Sci USA 103:3546–3551

Ríos G, Lossow A, Hertel B et al (2002) Rapid identification of Arabidopsis insertion mutants by non-radioactive detection of T-DNA tagged genes. Plant J 32:243–253

Rosso MG, Li Y, Strizhov N et al (2003) An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. Plant Mol Biol 53:247–259

Ryu CH H, You JH, Kang HG et al (2004) Generation of T-DNA tagging lines with a bidirectional gene trap vector and the establishment of an insertion-site database. Plant Mol Biol 54:489–502

Sallaud C, Gay C, Larmande P et al (2004) High throughput T-DNA insertion mutagenesis in rice: a first step towards in silico reverse genetics. Plant J 39:450–464
Samson F, Brunaud V, Balzergue S et al (2002) FLAGdb/FST: a database of mapped flanking insertion sites (FSTs) of Arabidopsis thaliana T-DNA transformants. Nucleic Acids Res 30:94–97
Sasaki A, Ashikari M, Ueguchi-Tanaka M et al (2002) Green revolution: a mutant gibberellin-synthesis gene in rice. Nature 416:701–702
Sasaki A, Itoh H, Gomi K et al (2003) Accumulation of phosphorylated repressor for gibberellin signaling in an F-box mutant. Science 299:1896–1898
Sasakuma T, Maan SS, Williams ND (1978) EMS-induced male-sterile mutants in euplasmic and alloplasmic common wheat. Crop Sci 18:850–853
Sato S, Nakamura Y, Kaneko T et al (1999) Complete structure of the chloroplast genome of Arabidopsis thaliana. DNA Res 6:283–290
Schmierer DA, Kandemir N, Kudrna DA et al (2005) Molecular marker-assisted selection for enhanced yield in malting barley. Mol Breed 14:463–473
Schwab R, Ossowski S, Riester M et al (2006) Highly specific gene silencing by artificial microRNAs in Arabidopsis. Plant Cell 18:1121–1133
Scott JW, Harbaugh BK (1989) Micro-Tom, a miniature dwarf tomato. Florida Agr Expt Sta Circ S-370
Sessions A, Burke E, Presting G et al (2002) A high-throughput Arabidopsis reverse genetics system. Plant Cell 14:2985–2994
Settles AM (2005) Maize community resources for forward and reverse genetics. Maydica 50:405–414
Settles AM, Holding DR, Tan B et al (2007) Sequence-indexed mutations in maize using the UniformMu trasposon-tagging population. BMC Genomics 8:116–128
Shibata D (2005) Genome sequencing and functional genomics approaches in tomato. J Gen Plant Pathol 71:1–7
Shikazono N, Suzuki C, Kitamura S et al (2005) Analysis of mutations induced by carbon ions in Arabidopsis thaliana. J Exp Bot 56:587–596
Shikazono N, Yokota Y, Tanaka A et al (1998) Molecular analysis of carbon ion-induced mutations in Arabidopsis thaliana. Genes Genet Syst 73:173–179
Shirley BW, Hanley S, Goodman HM (1992) Effects of ionizing radiation on a plant genome: análisis of two Arabidopsis transparent testa mutations. Plant Cell 4:333–347
Silverstone AL, Ciampaglio CN, Sun T (1998) The Arabidopsis *RGA* gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. Plant Cell 10:155–169
Slabbert R (2007) Development and evaluation of mutant germplasm of amaranthus tricolour. In: Ochatt S, Jain SM (eds) Breeding of neglected and under-utilized crops, spices and herbs. Science Publisher, Enfeild (NH), Jersey, Plymouth 137–150
Slade AJ, Fuerstenberg SI, Loeffler D et al (2005) A reverse genetic, non-transgenic approach to wheat crop improvement by TILLING. Nat Biotechnol 23:75–81
Somerville C, Dangl (2000) Genomics. Plant biology in 2010. Science 290:2077–2078
Song WY, Wang GL, Chen LL et al (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. Science 270:1804–1806
Soppe WJJ, Jacobsen SE, Alonso-Blanco C et al (2000) The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. Mol Cell 6:791–802
Spiegelman JI, Mindrinos MN, Frankhauser C et al (2000) Cloning of the *Arabidopsis RSF1* gene by using a mapping strategy based on high-density DNA arrays and denaturing high-performance liquid chromatography. Plant Cell 12:2485–2498
Steinmetz LM, Davis RW (2004) Maximizing the potential of functional genomics. Nat Rev Genet 5:190–201
Stern DB, Hanson MR, Barkan A (2004) Genetics and genomics of chloroplast biogenesis: maize as a model system. Trends Plant Sci 9:293–301
Straus D, Ausbel FM (1990) Genomic subtraction for cloning DNA corresponding to deletion mutations. Proc Natl Acad Sci USA 87:889–1893

Sun TP, Straus D, Ausbel FM (1992) Cloning Arabidopsis genes by genomic subtraction. In: Koncz C, Chua NH, Schell J (eds) Methods in Arabidopsis research. World Scientific, London, pp 331–341

Sundaresan V, Springer P, Volpe T et al (1995) Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. Genes Dev 9:1797–1810

Swarbreck D, Wilks C, Lamesch P et al (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Res 36:D1009–D1014

Tadege M, Ratet P, Mysore KS (2005) Insertional mutagenesis: a Swiss Army knife for functional genomics of *Medicago truncatula*. Trends Plant Sci 10:222–228

Takken FL, Schipper D, Nijkamp HL et al (1998) Identification and *Ds*-tagged isolation of a new gene at the *Cf*-4 locus of tomato involved in disease resistance to *Cladosporium fulvum* race 5. Plant J 14:401–411

Takumi S, Murai K, Mori N, Nakamura C (1999) Trans-activation of a maize Ds transposable element in transgenic wheat plants expressing the *Ac* transposase gene. Theor Appl Genet 98: 947–953

Tax FE, Vernon DM (2001) T-DNA-associated duplication/translocations in Arabidopsis. Implications for mutant analysis and functional genomics. Plant Physiol 126:1527–1538

Terada R, Johzuka-Hisatomi Y, Saitoh M et al (2007) Gene targeting by homologous recombination as a biotechnological tool for rice functional genomics. Plant Physiol 144:846–856

Terada R, Urawa H, Inagaki Y et al (2002) Efficient gene targeting by homologous recombination in rice. Nat Biotechnol 20:1030–1034

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408:796–815

Till BJ, Colbert T, Codomo C et al (2006) High-throughput TILLING for Arabidopsis. Methods Mol Biol 323:127–135

Till BJ, Reynolds SH, Greene EA et al (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. Genome Res 13:524–530

Till BJ, Reynolds SH, Weil C et al (2004) Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biol 4:12

Tissier AF, Marillonnet S, Klimyuk V et al (1999) Multiple independent defective suppressor-mutator transposon insertions in Arabidopsis: a tool for functional genomics. Plant Cell 11:1841–1852

Travella S, Klimm TE, Keller B (2006) RNA interference-based gene silencing as an efficient tool for functional genomics in hexaploid bread wheat. Plant Physiol 142:6–20

Uauy C, Distelfeld A, Fahima T et al (2006) A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. Science 314:1298–1301

Ueguchi-Tanaka M, Ashikari M, Nakajima M et al (2005) Gibberellin insensitive dwarf1 encodes a soluble receptor for gibberellin. Nature 437:693–698

Unseld M, Marienfeld JR, Brandt P et al (1997) The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides. Nat Genet 15:57–61

van der Biezen EA, Brandwagt BF, van Leeuwen W et al (1996) Identification and isolation of the *FEEBLY* gene from tomato by transposon tagging. Mol Gen Genet 251:267–280

Vizir IY, Mulligan BJ (1999) Genetics of gamma-irradiation-induced mutations in Arabidopsis thaliana: large chromosomal deletions can be rescued through the fertilization of diploid eggs. J Hered 90:412–417

Walbot V, Rudenko GN (2002) MuDR/Mu transposable elements in maize. In: Craig NL, Craigie M, Gellert AM, Lambowitz AM (eds) Mobile DNA II. ASM Press, Washington, DC, pp 533–564

Watanabe S, Mozoguchi T, Aoki K et al (2007) Ethylmethanesulfonate (EMS) mutagenesis of *Solanum lycopersicum* cv. Micro-Tom for large-scale mutant screens. Plant Biotechnol 24:33–38

Waterhouse PM, Helliwell CA (2003) Exploring plant genomes by RNA-induced gene silencing. Nat Rev Genet 4:29–38

Watson BS, Asirvatham VS, Wang LJ et al (2003) Mapping the proteome of barrel medic (*Medicago truncatula*). Plant Physiol 131:1104–1123

Weigel D, Glazebrook J (2002) Arabidopsis: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Weigel D, Ahn JH, Blázquez MA et al (2000) Activation tagging in Arabidopsis. Plant Physiol 122:1003–1013

Weil CF, Monde RA (2007) Getting the point-mutations in maize. Crop Sci 47:S60–S67

Wu GZ, Shi QM, Niu Y et al (2008) Shanghai RAPESEED Database: a resource for functional genomics studies of seed development and fatty acid metabolism of *Brassica*. Nucleic Acids Res 36:1044–1047

Wu JL, Fan YY, Li DB et al (2005) Chemical- and irradiation-induced mutants of *indica* rice IR64 for forward and reverse genetics. Plant Mol Biol 59:85–97

Wu SJ, Ding L, Zhu JK (1996) SOS1, a genetic locus essential for salt tolerance and potassium acquisition. Plant Cell 8:617–627

Würschum T, Groß-Hardt R, Laux T (2006) APETALA2 regulates the stem cell niche in the Arabidopsis shoot meristem. Plant Cell 18:295–307

Xu K, Xu X, Fukao T et al (2006) Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. Nature 442:705–708

Yasui T (2006) Waxy and low-amylose mutants of bread wheat (*Triticum aestivum* L.) and their starch, flour and grain properties. Jpn Agric Res Q 40:327–331

Yates JR 3rd (2004) Mass spectral analysis in proteomics. Annu Rev Biophys Biomol Struct 33:297–316

Yu J, Hu S, Wang J et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296:79–92

Yuan JS, Galbraith DW, Dai SY et al (2008) Plant systems biology comes of age. Trends Plant Sci 13:165–171

Yue SJ, Li H, Zhu YF et al (2007) Generation of transgenic wheat lines with altered expression levels of *1Dx5* high-molecular-weight glutenin subunit by RNA interference. J Cereal Sci 137:1253–1259

Zhang L, Fetch T, Nirmala J et al (2006a) *Rpr1*, a gene required for *Rpg1*-dependent resistance to stem rust in barley. Theor Appl Genet 113:847–855

Zhang SF, Ma CZ, Zhu JC et al (2006b) Genetic analysis of oil content in *Brassica napus* L. using mixed model of major gene and polygene. Acta Gen Sinica 33:171–180

Zhao Y, Thilmony R, Bender CL et al (2003) Virulence systems of *Pseudomonas syringae* pv. *tomato* promote bacterial speck disease in tomato by targeting the jasmonate signaling pathway. Plant J 36:485–499

Zhou B, Qu S, Liu G et al (2006) The eight amino-acid differences within three leucine-rich repeats between Pi2 and Piz-t resistance proteins determine the resistance specifity to *Magnaporthe grisea*. Mol Plant Microbe Interact 19:1216–1228

Chapter 19
Techniques in Plant Proteomics

Ludovít Škultéty, Maxym Danchenko, Anna Preťová, and Martin Hajduch

Abstract Plant proteomics is a relatively new research field that focuses on large-scale functional analysis of plant proteins. This new research field has already demonstrated immense potential for getting significantly deeper insight into the functional interaction of plant proteins and their roles in plant growth and development. The knowledge of key proteins responsible for valuable crop traits in the context of expression of other involved proteins plays pivotal role in discovery phase of crop improvement effort. This chapter introduces the field of plant proteomics in practical manner. Set of six figures helps to understand principles of protein extraction, separation and identification by mass spectrometry. Protein and peptide separation is introduced in greater details taking into account "gel-based" and "gel-free" methods.

19.1 Introduction

Proteins were mentioned for the first time in 1838 by Swedish chemist Jöns Jakob Berzelius. The name protein is derived from the Greek word "πρώτα"(prota) that could be translated to English as "of primary importance". Proteins are functional molecules that operate through the all metabolic and regulatory pathways in the cells, tissues and organisms and thus proteins are key players in improvement of nutritional and agricultural traits of crop plants. Proteins are often organized into multiprotein complexes that function as molecular drivers of many processes and are localized within specialized subcellular compartments. The world of proteins is very complex. Due to alternative transcription initiation and/or alternative splicing, the expression of single gene might produce several transcripts. One transcript is most likely to be translated to more than one protein due to alternative translation

L. Škultéty
Institute of Virology, Centre of Molecular Medicine, Slovak Academy of Sciences, Bratislava, Slovakia

M. Danchenko, A. Preťová and M. Hajduch (✉)
Institute of Plant Genetics and Biotechnology, Slovak Academy of Sciences, Nitra, Slovakia
e-mail: hajduch@savba.sk

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_19, © Springer Science+Business Media B.V. 2010

initiation. Additionally, these proteins might be further posttranslationaly modified (for instance reversible phosphorylation temporary "turn off" the protein activity). With such high level of complexity, large scale protein analysis (proteomics) becomes necessary tool of advanced plant science research. The word "proteomics" was first time mentioned in 1996 by Marc Wilkins, who was that time graduate student at Macquarie University in Australia (Wilkins et al. 1996) as protein alternative to genomics. Proteomics can be called a post-genomics methodology, because the success of protein identification greatly relies on completion of genomic sequencing and characterization of individual proteins. Proteomics aims at the analysis of entire proteome, all proteins within a particular biological sample. Such information will provide comprehensive knowledge regarding metabolic processes that occurs in the investigated system. However, this ambitious goal is not feasible at present because of extremly broad range in abundance of protein species and detection limits of current mass spectrometers. Despite development of mass spectrometers toward detection of single molecule, presently only small percentage of entire proteome is possible to unravel. The ambition of this chapter is to provide aid to any researchers considering or initiating plant proteomics experiments.

19.2 Protein Extractions

Each proteomic approach starts with a sample preparation. Thus, an efficient protocol is absolutely essential and must be developed to release the proteins in a soluble form. There is no 'perfect protocol' and the optimal procedure must be determined individually for each sample type. Ideally, a protein extract should contain all proteins from the sample with no contamination of other compounds.

An extraction of proteins from plant tissues is challenging. Plant extracts are rich in proteases, oxidative enzymes, and compounds that interfere with subsequent protein separation and identification. These include cell wall polyphenols, polysaccharides, starches, oils, and various secondary metabolites. Salts, chlorophyll and other molecules (lignins, pigments, etc.) that are contained in the differentiated green tissues at high levels might also interfere with proteomic analysis. In addition, a relatively large abundance of photosynthetic proteins in leaves and storage proteins in seeds dominating protein profiles prevents detection of low abundant proteins (Isaacson et al. 2006).

In order to isolate the proteins of interest, the specific plant tissues must be disrupted in the first step. Several disruption techniques, both mechanical and chemical are available. Usually, the sample is grounded in liquid nitrogen using amortar and apestle. Upon cell disruption proteases may be liberated, that could greatly complicate the protein separation. Thus protein sample should be protected from unwanted proteolysis by addition of a cocktail of five to six protease inhibitors and keeping the protein mixture cooled during the preparation. Choosing an appropriate protocol largely depends on the nature of the extracted tissue and on the downstream application (e.g., analysis of protein complexes, membrane proteins or

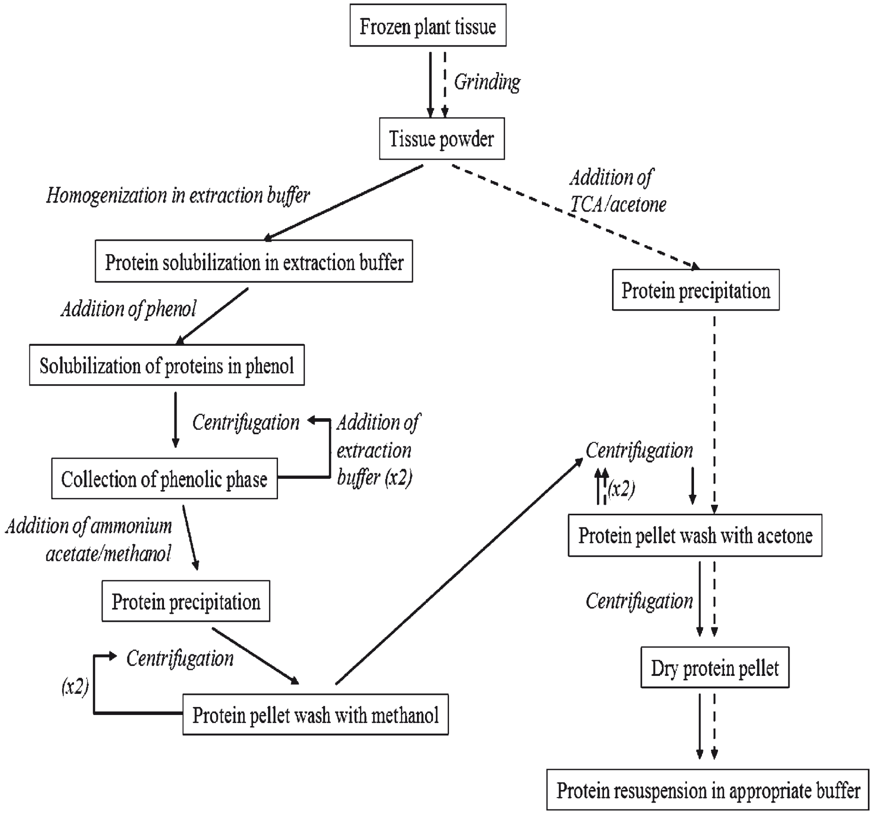


Fig. 19.1 Flowchart of protein extraction protocol. Solid lines indicate phenol extraction method; dashed lines indicate TCA/acetone extraction method

proteins with specific post-translational modifications). So far, there are two dominant methods for protein isolation from plant tissues (Fig. 19.1). The most common method is based on precipitation of proteins from homogenized tissue or cells with trichloroacetic acid (TCA) in acetone. This protocol has successfully been used with tissues from young plants; however, it was not a best choice for more complex plant tissues. An alternative protocol is based on the solubilization of proteins in phenol, followed by their precipitation with ammonium acetate in methanol. This method was developed in the 1990s and was used successfully with olive leaves, wood, and recalcitrant tissues such as orange pee, tomato, banana and avocado fruits. Although the phenol-based method is more time consuming and laborious than the TCA precipitation, it has the potential to generate samples of high purity. The compounds such as polysaccharides and other water-soluble contaminants are partitioned into a discrete aqueous phase or are centrifuged into a separate pellet, away from the protein-enriched phenolic layer. Additionally, the combination of phenol and TCA/acetone precipitation was proposed (Wang et al. 2006).

Table 19.1 The summary of basic extraction methods used for various plant tissues

Plant	Extraction method	Reference
Soybean seed	Phenol	Hajduch et al. (2005)
Brassica seeds	Phenol	Hajduch et al. (2006)
Soybean seeds	Urea, urea/thiourea, phenol, TCA/acetone	Natarajan et al. (2005)
Soybean cotyledons	TCA/acetone	Narajan et al. (2005)
Cultured root tissues	TCA/acetone	Narajan et al. (2005)
Tobacco flowers	TCA/acetone	Narajan et al. (2005)
Apple fruit	Urea, TCA/acetone, hot SDS	Song et al. (2006)
Banana	SDS lysis, phenol	Song et al. (2006)
Potato tubers	Phenol/TCA	Delaplace et al. (2006)
Recal citrant plant Ttissues	Phenol	Wang et al. (2006)
Tomato fruit	Phenol	Faurobert et al. (2007)

The procedure is universally applicable for high efficient extraction of proteins from plants tested on a wide range of tissues. The protocol was evaluated on different problematic tissues high in polyphenols, sugars or pigments; low in proteins or (with) extreme acidity. In all cases acceptable protein quality profile was achieved.

Overall, a great deal of strategy and consideration are required in choosing the right sample preparation protocol. Up to date, number of methods has been published (Canas et al. 2007). Several examples of methods for protein extraction from various plant tissues are summarized in the Table 19.1.

19.3 Protein Separation

Separating and analyzing each tiny poppy seed inside a poppy head would be an extremely difficult task. Reduce these poppy seeds to sub-microscopic size and you'll begin to understand the challenge of proteomics, which main goal is to analyze complete proteome. Although new technologies have been developed, separation of one protein from others is typically the most laborious and limiting aspect of proteomics. Thus, reliable and effective methods of sample preparation and separation are keys to success of proteomic research.

Crude protein extracts contain not only proteins of interest but also interfering substances, such as salts, small ionic molecules, ionic detergents, saccharides, lipids, and other non-protein components. It is necessary to deplete or entirely remove these interfering substances from the sample. Their presence in the sample may lead to difficulty in further protein separation and also disturb the detection and identification of proteins in proteome studies. Depending on the type of sample, there are various ways to prepare protein sample for further separation (Zellner et al. 2005). The basic methods include precipitations, dialysis, ultrafiltration and ultracentrifugation. In bulk protein purification, a common first step to isolate proteins is precipitation relying on different chemical principles. It can be performed by ammonium sulfate, TCA,

TCA in acetone, ethanol, or acetone (Jiang et al. 2004). Although many protein precipitation methods have the advantages for concentrating and eliminating interferences, they also have the disadvantages of protein irreversible denaturation and protein insolubilization. Dialysis is an old established procedure for reducing the salt concentration in sample and its separation based on principles of diffusion that allows the low molecular weight contaminant removal from sample solutions. Using dialysis, the maximal interfering substances can be reduced. However, the proteins in sample might be lost, get the high volume of interchanged buffer and spend more time than with other desalting techniques (Manza et al. 2005). Ultrafiltration and ultracentrifugation can remove high or low molecular weight interfering substances in a relatively short time by using either selective permeable membranes (Tirumalai et al. 2003; Haper et al. 2004) or a centrifugal force (Lebowitz et al. 2002), respectively. Although the removal of these interferences can concentrate proteins, some with similar molecular weight as proteins are also concentrated. It means that each purification method has some advantages and disadvantages in the same time. Therefore, an alternative or multi-steps method might be required for protein concentration and/or desalting in order to get sample of high quality.

Plant proteome is highly complex and consist of thousands of protein species with different chemical and physical properties. Because of limited resolution power of analytical separation techniques, usually only most abundant proteins are identified by subsequent mass spectrometry analysis (MS). Since abundance of protein species may differ by 7–10 orders of magnitude, the relatively low abundant proteins are usually masked by more abundant ones. This makes difficult to relate results of proteome profiling to the system biology. Low copy number regulatory proteins such as kinases, phosphatases, or GTPases can be detected only after applying additional fractionation strategies to reduce sample complexity (Stasyk and Huber 2004).

To gain a better understanding of plant metabolism, initial pre-fractionation methods can be used. These include protein and peptide affinity purification (Lee and Lee 2004), protein chromatography (Lescuyer et al. 2004), preparative protein isoelectric focusing (Görg et al. 2002) and fractional centrifugation or other sample treatment methods (Righetti et al. 2005).

Even initial pre-fractionation techniques are applied, the protein sample is still far too complex to be directly analyzed by MS for protein identification. Depends on further techniques of complexity reduction, either “gel-based” or “gel-free” proteomics approaches are used.

19.3.1 Gel-based Proteomics

One dimensional gel electrophoresis (1-DE) is a common laboratory technique that can be use both as preparative and analytical method. The principle of electrophoresis relies on the movement of a charged ions in an electric field. In practice, the proteins are usually denatured in a solution containing a detergent – sodium

dodecyl sulphate (SDS). In these conditions, protein to protein interactions are prevented, proteins are unfolded and negatively loaded SDS-protein complexes are formed. The amount of detergent bound is so large that any differences in native charge are eliminated. The bigger the molecule, the more SDS is bound, so that all macromolecules treated with SDS have the same charge to mass ratio. Thus, the force per unit mass in an electric field is the same, and all molecules should have the same velocity if there is no frictional drag. But electrophoresis with SDS is nearly always carried in gels that suppress connective currents produced by small temperature gradients and serves as molecular sieves. During SDS gel electrophoresis, the SDS-protein complex moves in electric field toward the positive pole. Molecules that are small compared with the pores in the gel readily move through the gel, whereas molecules much larger than the pores are almost immobile. Intermediate-size molecules move through the gel with various degrees of facility. Polyacrylamide gels are choice supporting media for electrophoresis because they are chemically inert and are readily formed by the polymerization of acrylamide. Moreover, their pore sizes can be controlled by choosing various concentrations of acrylamide and a cross-linking reagent (piperazine, methylenebisacrylamide) at the time of polymerization.

However, it became quickly apparent that SDS-PAGE is not able to separate complex protein mixtures. Thus, two dimensional gel electrophoresis (2-DE) was introduced in 1975 (O'Farrell 1975; Klose 1975). This technique sorts proteins according to two independent properties in two discrete steps: the first-dimension step – isoelectric focusing (IEF) separates proteins according to their isoelectric point (pI); the second-dimension step - SDS-PAGE separates proteins according to their molecular weight (MW) (Fig. 19.2). The result looks like a Dalmatian's coat with spots corresponding usually to a single protein species. Thousands of different proteins and their isoforms can thus be separated and information such as the protein pI, the apparent MW, and the amount of each protein can be determined. However, for a long time 2-DE led a limited existence due to gel to gel variation. This status improved by introduction of immobilized IEF gradient strips that dramatically increased reproducibility as well as the resolution of 2-DE by an order of magnitude (Corbett et al. 1994). The presence of an immobilized pH gradient inside the IPG strip has a crucial importance. Under the influence of an electric field, a protein as an amphoteric molecule will move to the position in the pH gradient where its net charge is zero. A protein with a negative net charge will migrate toward the anode, becoming less negatively charged until it reaches zero net charge. A positively charged protein moves at the same time through the pH gradient until it also reaches its pI. If a protein is going to diffuse away from its pI, it immediately gains charge and migrates back. This is the focusing effect of IEF, which concentrates proteins at their pIs and allows proteins to be separated on the basis of very small charge differences. After focusing, the proteins in the IPG strip have to be denatured, reduced and alkylated. Each strip is then transfered onto SDS-PAGE gel and the electrophoresis is carried out (Rabilloud 2002; Rehm 2006).

Several detection methods are widely used for on-gel protein visualization after SDS-PAGE or 2-DE. Most popular, colloidal Coomassie blue staining

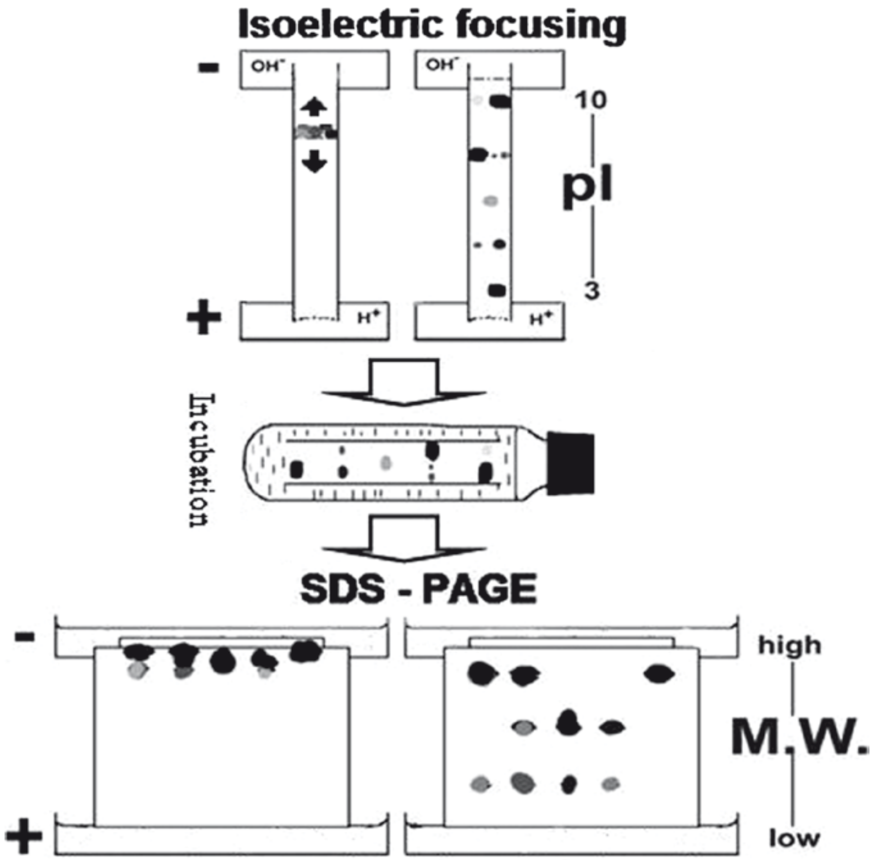


Fig. 19.2 Schematic diagram of two-dimensional electrophoresis. The technique sort proteins according to two independent properties in two discrete steps: the first-dimension step – isoelectric focusing separates proteins according to their isoelectric point (pI); the second-dimension step – polyacrylamide gel electrophoresis in presence of SDS separates proteins according to their MWs

(Neuhoff et al. 1988) provides good dynamic range of detected proteins and excellent compatibility with MS. It has gained popularity in the early days of proteomics, when its sensitivity matched almost exactly the needs in protein amounts required for a decent analysis by MS. Presently when more sensitive MS instruments are available, detection by colloidal Coommassie causes bigger consumption of sample to gain comparable number of detected proteins in comparission to more sensitive stains. However, if the quantity of sample is no problem, it is still method of choice due to very easy and cheep implementation, and great linear dynamic range. Fluorescent detection methods offer a good alternative, and thus the metal chelate-based methods (Berggren et al. 2000; Rabilloud et al. 2001) have become increasingly popular. More recently, these methods offer an interesting sensitivity allied to a very good compatibility with MS analysis (Lamanda et al. 2004).

Last but not least, silver staining still offers the maximal sensitivity, and all the ancillary advantages associated on light absorption-based methods, such as easy visualization, spot excision, and quantitation despite poor linear dynamic range. Development of silver-ammonia (Chevallet et al. 2006) and other methods (Shevchenko et al. 1996) solved poor compatibility of silver staining with MS. The ammoniacal silver is good choice for silver-based detection methods for proteomics studies. However, it must be kept in mind that ammoniacal silver is not compatible with all electrophoresis systems. Although it is fully compatible with the popular Tris-glycine system and the more recent Tris taurine system (Tastet et al. 2003), it is not compatible with the Tris-tricine systems.

To take full advantage of the 2-DE separation, multiplex quantitative analysis of the component proteins of related but different protein samples can be performed on the same gel. This technique is known as two dimensional difference in-gel electrophoresis (2D-DIGE) and allows labeling protein mixtures with fluorescent cyanine dyes, such as Cy2, Cy3 and Cy5 maleimides (Alban et al. 2003). These CyDyes are structurally similar, but spectrally different (Cy2, λem = 520 nm; Cy3, λem = 580 nm; Cy5, λem = 670 nm). Compared to conventional 2-DE, 2D-DIGE has the major advantage because control and experimental sample are run in the same gel. These samples are then scanned using different wave lengths and because they were run in the same gel, the images can be perfectly overlaid (Fig. 19.3). This reduces the number of gels that must be run to make statistically valid comparisons and raises the confidence with which protein changes between samples can be detected and quantified. Use of a third dye (Cy2) permits an internal standard to be created by pooling an equal aliquot of all biological samples in the experiment. The internal standard is then run on every gel in the experiment. This means that every protein spot from all samples will be represented in the internal standard. This in turn allows more accurate quantification and spot statistics between gels.

In addition, the post-translational modifications of phosphoproteins and glycoproteins can be detected directly on the gel (Sriyam et al. 2007). A new phosphoprotein specific fluorescence-dye called "Pro-Q™ Diamond" recently became available from Molecular Probes (Eugene, OR, USA), and can be used to detect phosphorylated tyrosine, serine, or threonine residues of proteins on SDS-PAGE and 2-DE (Steinberg et al. 2003; Agrawal and Thelen 2005). Additionally, a recently developed approach for the detection of glycosylated proteins relies upon the utilization of a fluorescent hydrazide. Pro-Q™ Emerald 300 and Pro-Q™ Emerald 488 glycoprotein stains provide an attractive alternative to the labeling with radioactive sugars that are conjugated to glycoprotein by periodic acid Schiff's mechanism (Wu et al. 2005). Gels stained with both Pro-Q™ Diamond phosphoprotein stain and Pro-Q™ Emerald glycoprotein stain can be also post-stained with SYPRO Ruby dye, which allows sequential detection of total protein profile in the same gel.

Evaluation of only two visualized high-resolution 2DE gels by manual comparison is not an easy task. In large studies with patterns containing several hundreds or thousand spots, it may be almost impossible to detect the appearance of a few new spots or the disappearance of single spot. Image collection hardware and image evalu-

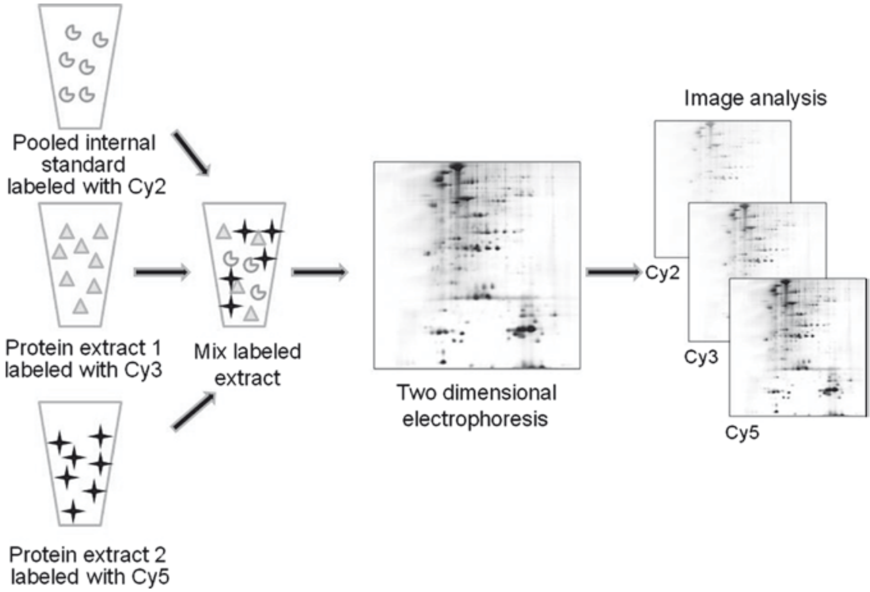


Fig. 19.3 Schematic diagram of the DIGE technology platform. Two different samples are derivatized with two different fluorophores (Cy3 and Cy5), combined and then run on a single 2-D gel together with internal standard labeled with Cy2. Proteins are detected using a dual laser scanning device or xenon-arc-based instrument equipped with different excitation/emission filters in order to generate three separate images. The images are then matched by a computer-assisted overlay method, signals are normalized, and spots are quantified

ation software are necessary to detect these differences as well as to obtain maximum information from the gel patterns. The systems that are commonly available in retail (ImageMaster 2D Platinum or DeCyder from GE, PDQuest from BioRad, Melanie from GeneBio, Dymension from Syngene, Delta 2D from Decodon, Progenesis from Nonlinear Dynamic, etc.) seamlessly perform classical or DIGE gel analysis. These systems provide powerful solutions from data acquisition to protein spot information. The spot detection and matching algorithms facilitate the extraction of statistically valid differences between groups of 2-D gels, while requiring minimal user intervention. The application integrates filtering, querying, reporting, statistical and graphing options so it is easily to view, compare, analyze and present the results.

As analytical 2-DE is remarkably well suited to studying protein expression in biological systems, it is used in most laboratories where proteomics are performed. However, this separation technique has some drawbacks that can severely limit the ability to monitor protein expression on a truly global scale. Proteins notably difficult to separate using 2-DE are membrane, low copy number, large (>150 kDa), and highly basic proteins. Driven by the urgent need to study these “difficult” proteins together with their less challenging counterparts, some alternative separation methods have to be used.

19.3.2 Gel-Free Proteomics

Although, the methods that performs protein/peptide separations in the liquid phase generally have a lower ability to resolve complex mixtures and provide inferior visualization of proteomes as compared to 2-DE, their advantages are analysis of difficult proteins, flexibility, relative speed, and ease of sample handling automation. Because handling of samples occurs in the liquid phase, it is feasible to automate a large part of the experimental procedures, including sample preparation for MS, data acquisition, and database searches. This automation also shortens the time between sample preparation and protein identification. Liquid separations can be performed using chromatography, electrophoresis, or a combination of these methods.

Chromatography is usually used to separate different compounds in a mixture and to determine the exact amount of each compound. Many types of available matrix used for column chromatography are usually packed in the column in the form of small beads. Because some of the chromatographic separations are non-denaturing, intact proteins can easily be mass-measured prior to enzymatic or chemical degradation. Also, the lack of free acrylamide in the separation process prevents any chemical blocking of the amino terminus of the separated proteins.

Reversed-phase liquid chromatography (RP-LC) is the most used method for peptide separation. Additionally, it can efficiently separate also small stable proteins. Peptide and/or proteins adsorb to the hydrophobic surface that consists of porous silica particles coated in general with n-alkyl chains (Fig. 19.4a). For peptides, silica particles with pore dimensions of 100 to 300 Å are used. For proteins, the pore diameter should be ten times bigger. The n-alkyl chains are 4-, 8-, or 18-C long and their length unpredictably changes the separating properties of the RP-LC. Two peptides that show two separate peaks on C18 columns may exhibit only one peak on C4 columns (or vice versa). The sample is normally eluded with a rising acetonitrile gradient. Methanol or 2-propanol also serves well. The quality of separation depends beside of the beads also on the steepness of the gradient and the temperature. The temperature is in play because peptide can maintain a secondary structure (α-helix, ß-fold), which influences the adsorption (high temperature prevent secondary structures). Regarding the column dimension (typically 10–20 cm), the separation of peptides and smaller proteins improves with longer columns. On the other hand larger proteins are separated on shorter columns because the yield becomes too low. However, RP-LC is rarely used to separate larger proteins because they like to denature under these conditions (Rehm 2006).

Many soluble proteins have hydrophobic areas on their surface. These areas might be associated with hydrophobic surfaces in aqueous solution. To increase these interactions, high concentrations of certain ions (NH4+, Rb+, K+, Na+, PO_4^{3-}, SO_4^{2-} etc.) should be added. Whereas chaotropic salt disrupts the water structure and thereby reduces the inclination for hydrophobic interactions.

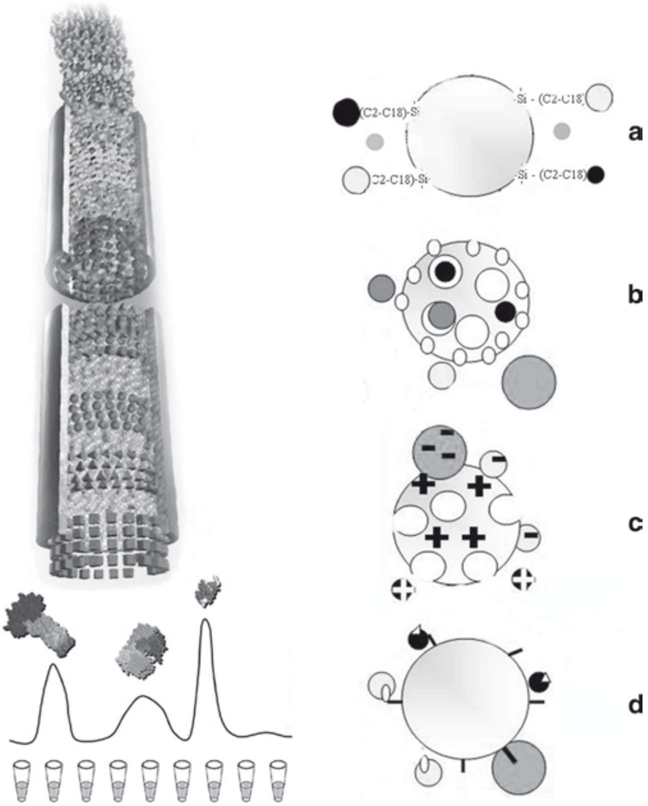


Fig. 19.4 Schematic diagram of chromatographic separations. a: Reverse phase chromatography. Peptide and/or proteins adsorb to the hydrophobic surface that consists of porous silica particles coated in general with n-alkyl chains. The chains are 4-, 8-, or 18-C long. The sample is normally eluded with a rising acetonitrile gradient. b: Size-exclussion chromatography. The columns separate proteins according to their size. The matrix consists of tiny porous beads. Protein molecules that are small enough to enter the holes in the beads are delayed and travel more slowly through the column. c: Ion-exchange chromatography. The column is packed with small beads that carry positive or negative charges that retard proteins of the opposite charge. The association between a protein and the matrix depends on the pH and ionic strength of the solution passing down the column. d: Affinity chromatography. The columns contain a matrix covalently coupled to a molecule that interacts specifically with the protein of interest (e.g., antibody or an enzyme substrate). Protein that binds specifically to column can be released by pH salt gradient

The protein sample is loaded with high ionic strength (largely ammonium sulfate) onto a hydrophobic matrix and the captured proteins are eluded with detergents that cover the hydrophobic areas of the proteins and thereby loosen them from the matrix. Matrices for hydrophobic chromatography are derivatized either with

phenyl residues (e.g., phenylsepharose) or octyl residues (e.g., octylsepharose) (Wu and Karger 1996).

Size-exclusion chromatography (SEC), also known as gel permeation chromatography or gel-filtration chromatography, is generally considered the premier method for determining the molar mass averages of polydisperse macromolecules, separation of target protein in a mixture, and salt removal or buffer exchange. The dissolved analyte is injected onto a column packed with porous, inert material (porous gels or other rigid inorganic packing particles) and is carried through the column by solvent (Fig. 19.4b). Both molecular weight and shapes of proteins contribute to the degree of retention. Smaller molecules diffuse further into the pores of the beads and therefore move through the bed more slowly, while larger molecules enter less or not at all and thus move through the bed more quickly. The process has been described as an inverse molecular sieving mechanism that depends on hydrodynamic volume of a dissolved molecule with respect to the average pore size of the column packing material (Striegel 2008). Although it is widely used, low-pressure SEC is not ideal for protein purification for several reasons: (1) the resolution is bad, (2) the sample volume is limited, and (3) the chromatography takes a long time because the flow rate is limited. In addition, (4) the sample is diluted by at least a factor of 3.

In ion exchanger chromatography (IEC), proteins bind to a matrix via electrostatic interactions. The matrix carries positively charged groups (anion exchanger) or negatively charged groups (cation exchanger). Extent and strength of the binding of a protein to the ion exchanger depends on the pH and ion strength of the buffer, the pI of the protein, and the density of the charges on the matrix (Fig. 19.4c). Technically speaking, the IEC is easier than the SEC. The IEC does not require perfectly poured column, and the sample volume can be larger than the column volume. The charged sample is loaded on the column which is equilibrated with a buffer of low ion strength (e.g., 20 mM salt). Unbound proteins are washed out and the captured protein is eluded through increasing the salt concentration or changing the pH. The easiest way is to chromatograph at a steady pH. The captured protein is then eluded through stepwise changes in ion strength. With bigger proteins and integral membrane proteins, it is fruitless to try to improve the IEC with a salt or pH gradient. The concentration effect of the IEC gets lost, and the purification factors are increased only slightly. It is more efficient and reproducible to elude the protein with a well-balanced step gradient (Choudhary and Horvath 1996).

Selective separation of a specific protein or group of proteins can be achieved using affinity chromatography that is based on the ability of a biologically active molecule to bind specifically and reversibly to a complementary molecule. The binding sites of the immobilized substances must be sterically accessible after their coupling to the solid support and should not be deformed by immobilization. In the case of specific proteins, an affinant is attached to the active surface of the column packing material or column surface. The sample is injected onto the column and the protein(s) of interest captured by the affinant (Fig. 19.4d). Compounds that do not possess a complementary binding site for the bound ligand will either

pass directly through the column or be eluted by a low-stringency washing step. The bound protein(s) is then recovered by washing the column with a competitive substrate or a solution that disrupts the interaction between protein and affinant (e.g., denaturants). While the use of antibodies directed to a specific protein remains the most popular affinity-based fractionation method, many other affinity techniques to isolate a specific class of protein or peptide have been developed. These methods include immobilized metal ion affinity chromatography (IMAC) containing nickel or copper ions, to capture histidine-containing peptides (Wang et al. 2002; Tishchenko et al. 2002) or gallium and alternatively zirconium to isolate phosphopeptides (Feng et al. 2007; Posewitz and Tempst 1999). In addition, affinity methods have been developed to select peptides containing specific types of residues such as cysteine, tryptophan or methionine (Zhang et al. 2004). There are as well a variety of different lectins that have been used to selectively separate glycoproteins based on the composition of the carbohydrate side chain (Madera et al. 2005).

Since the resolving power of a single chromatographic step is very limited and because of use of different types of columns and solvent systems naturally leads to different protein or peptide separation, efforts have been made to develop multidimensional approaches. Recently, a separation of peptides by 2-dimensional liquid chromatography in conjunction with MS was commercially introduced under the name MudPIT (Multi Dimensional Protein Identification Technology (Wolters et al. 2001; Wang and Hanash 2003). The first dimension is normally a strong cation exchange column (SCX), as these have high loading capacities. The second dimension is RP-LC, which complements the SCX as it is efficient at removing salts and has the added advantage of being compatible with electrospray (ESI) MS analysis. The sample preparation is relatively straightforward, the samples are denatured, the cysteines reduced and alkylated, and the proteins digested with a protease such as trypsin. The samples are then acidified and loaded onto the SCX column. Charged peptides bind to the SCX column, whereas any uncharged peptides pass through and bind to a reverse phase trap column. The peptides are then eluted from the trap column onto an analytical RP column, using a reverse phase gradient, separated and eluted into a tandem mass spectrometer. Peptide fragmentation data is then obtained to identify the peptides and hence the proteins from which they are derived. In the next step, salt at a particular concentration is injected onto the SCX column, displacing further peptides from it onto the RP trap column. Salt is removed by washing, an analytical RP separation is performed, and the eluting peptides are analyzed by MS. Incremental increases of salt are used. The end result is multiple protein identifications from each salt step (Fig. 19.5).

An advantage of performing a liquid-phase separation is the inherent flexibility available for experimental design. It is not necessary to limit protein separation to chromatographic or other approaches; any combination may be tried and found to be especially useful for a given biological system (Nilsson and Davidsson 2000; Issaq et al. 2002).

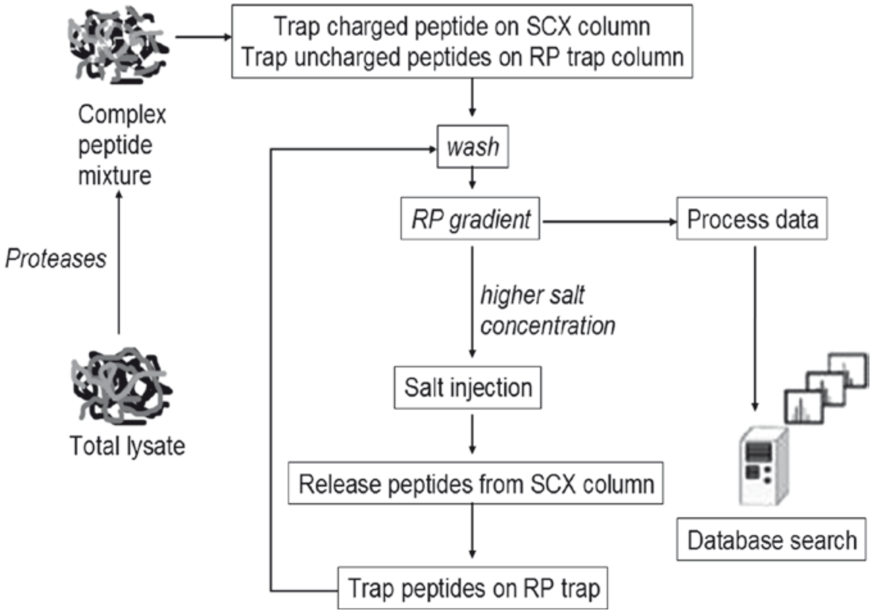


Fig. 19.5 Schematic diagram of the MudPIT approach. The protein mixture is divided in aliquots, which are each digested by a protease. After protease inactivation, the aliquots are pooled and acidified. This procedure is intended to maximize the probability for every protein to be represented by at least one peptide in the subsequent analysis. The peptide mixture is then loaded on a strong cation exchanger (SCX) column, followed by a reverse phase column (RP) coupled by a nanospray device to a tandem mass spectrometer. The SCX column is eluted by salt gradient. This transfers a population of peptides in the RP column, where they bind. Using a reverse phase gradient, the separated peptides are eluted into a tandem mass spectrometer. Peptide fragmentation data is then obtained to identify the peptides and hence the proteins from which they are derived. When the reverse phase gradient is fully developed, a step of more concentrated salt is performed on the SCX column, etc

19.3.3 Alternative Separation Technologies

The Rotofor cell (Bio-Rad) has been developed for preparative scale IEF in liquid phase (Dolnik 2008; Davidssona and Sjogrena 2005). This technique has a unique ability to enrich low-abundance proteins up to 500 times at their respective isoelectric points. The individual proteins might be then isolated on the basis of their size differences in the liquid phase continuous SDS-PAGE (Pons et al. 2005; Rathinasabapathi et al. 2001). The Prep cell (Bio-Rad) has been designed for that purpose. Two-dimensional preparative liquid-phase electrophoresis (2D-LPE) allows high protein loads (up to 1 g) and large volumes (up to 55 mL), thus yielding sufficient amounts of low abundance proteins for further characterization by MS (Yuana and Desiderioa 2005).

A novel approach similar to 2-DE is under development at Lynx Therapeutics (Hayward, California). As in 2-DE systems, the Lynx Protein Profiler uses separation by charge combined with separation by mass. However, the gels are replaced by flat plates that contain multiple channels. In the first-dimension, electrophoresis in an entangled polymer sieving solution is performed for separation based upon mass. In this system, proteins are not complexed with SDS but are fluorescently prelabeled for detection. The separation range is 6,000–200,000 MW. After the first-dimension separation, the separated proteins are electrically driven orthogonally into 100 parallel channels, which are coated with covalently attached buffers. Applying an electric field to these channels generates a stable pH gradient within which proteins are resolved upon the basis of their isoelectric points (Wehr 2001).

An alternative approach for complex mixture analysis has been developed also by Ciphergen Biosystems, Inc. (Fremont, California). This technology, marketed as the Protein-Chip system, selectively captures the proteins of interest using an aluminum strip with eight spots that carry different affinity adsorbents. The adsorbents might be nonspecific, such as cationic, anionic, hydrophobic, or hydrophilic materials, or highly specific, such as antibodies or receptors. Proteins are adsorbed from a complex mixture, unbound material and interferences are washed away, and bound species are analyzed by MS systems (Chapman 2002; Vorderwülbecke et al. 2005).

In addition, microscale lab-on-a-chip devices have been developed for performing chemical reactions and separations (Mouradian 2002; Peng et al. 2008; Le Nel et al. 2008). Several versions of this technology have been commercialized, but all rely on similar principles of microfluidics. Reagents and chemicals are transported electrically using electroosmotic flow or, in some cases, by hydrodynamic flow using pressure or vacuum. Analytes are separated by electromigration using electrophoresis and electroosmosis.

19.4 Protein Identification

Two methods, Edman degradation and MS, are used for protein identification. A more powerful method is MS that determine the precise molecular weights of specific cleavage peptides and/or fragments derived from a given protein. These data are compared with theoretical protein sequences calculated from the protein sequence databases, that in principle results in protein identification. In addition, MS could be used for identification and location of post-translational protein modifications.

19.4.1 Protein Digestion

Protein identification is based on the knowledge that endoproteases cleave protein at certain sides (Adams et al. 1999). Although hundreds of endoproteases are

known, only a small number can be used in protein analyses. They needs to be available in large quantities, high purity, and well characterized. Among of those that fulfill these criteria, trypsin is most widely used. This proteolytic enzyme specifically hydrolyzes proteins at the C-terminal side of arginine and lysine, except those when following C-terminal residue is proline.

When proteins are separated either with one or two dimensional electrophoresis, "in-gel" digestion should be performed prior MS analysis (Jensen et al. 1999). Generally, excised gel pieces are destained, reduced, and alkylated prior digestion with trypsin. The peptides are then eluted from the gel. "In-solution" digestion is used for "gel-free" proteomics, where the proteins are cleaved directly in the liquid phase before separation (Wolters et al. 2001).

19.4.2 Mass Spectrometry

Mass spectrometry can, in principle, sequence any size of protein, but the problem becomes computationally more difficult as the size increases. Thus, usually peptides obtained by cleavage of the protein are used for identification. There are two basic methodologies for protein identification - peptide mass fingerprinting (PMF) and tandem mass spectrometry (MS/MS). While PMF provides exact mass-to-charge ratio (m/z) of ions derived from a peptide mixture, MS/MS approach goes further and fragment these precursor (parent) ions and thus provides also peptide sequencing information (Fig. 19.6). In addition, when database search did not result in positive match or no comprehensive database is available for studied organism, peptide fragmentation data can be used for direct deduction of peptide sequence (DeNovo sequencing).

The mass spectrometer in general consists of three distinct components: the ionization source, the mass analyzer, and the detector. The ionization source is the region of the instrument in which the sample of interest is ionized, with a positive or negative charge. The mass analyzer is where ions created in the source region are guided through the instrument to the detector, where their m/z ration is measured. Two soft ionization sources-electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) (Karas et al. 1985) are currently the principal methods for peptide/protein ionization in proteomics. MALDI coupled with time-of-flight (TOF), known as MALDI-TOF, is used predominantly for the analysis of simple peptide mixtures by means of PMF. For the MALDI TOF analysis, sample is co-crystallized with a matrix on MALDI plate. The matrix is typically a small organic molecule that has two functions: (1) help to crystallize a sample and (2) absorb light at specific wavelength. The mixture is excited by a pulsed laser. The matrix absorbs the laser light and the mixture ablates from the surface. The sample molecules are surrounded by matrix as they leave the surface. After a short distance the sample begins to "desolvate." During the desolvation process a proton is transferred from the matrix to the analyte forming the MH+ or MH_2^{2+} ions which are

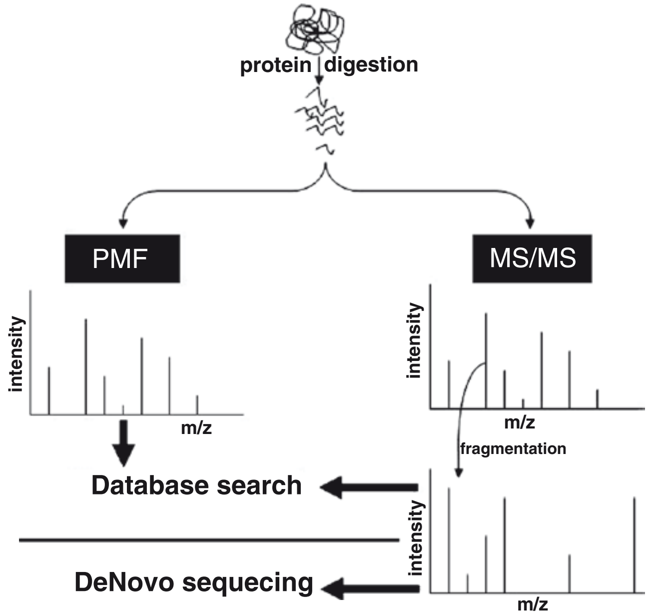


Fig. 19.6 Protein identification strategy. After protein cleavage with protease (usually with trypsin), resulted peptides are analyzed either by peptide mass fingerprinting (PMF) or tandem mass spectrometry (MS/MS). Peptide mass fingerprinting provides a list of m/z values of peptides for database search, while MS/MS approach further fragment selected peptides and m/z values of peptide fragments can be use either from database searches or for direct deduction of peptide sequence (DeNovo sequencing)

typically measured in TOF analyzer. This analyzer works on simple principle: time of flight for particular ion is proportional to its m/z value. Although, MALDI-TOF instrument is sensitive, robust, easy-to-use and is able to participate in automatic flow-through, it has a limitation in generation of PMF data only. The combinations of MS technologies, such as TOF/TOF or hybrid quadrupole time-of-flight (Q-TOF) are even more advantageous for protein identification.

Nowadays, an MS/MS coupled to a liquid chromatography (LC-MS/MS) represents the most widely used proteomic technology applied for protein identification. It relies on separation of peptides by LC prior the analysis on MS/MS instrument. A solution containing peptide is usually pushed through a capillary which is charged by high voltage. As the liquid leaves the capillary droplets are formed and analyte can pick up some of the charge as it is desolvated (Alexandrov et al. 1984). This ionization is called electrospray (ESI), or nanospray (nls/min flow rate) ionizations (NSI) that is in principle directly impacted by the solution phase chemistry of peptides that varies in accordance to their physicochemical properties, including pKa value, polarity, hydrophobicity, and ionization potential, and by the concentration and type of peptides infused into the ionization source. Produced ions are then introduced into the tandem mass analyzer and m/z values are acquired. Since the

droplets may contain many protons, the analyte could be highly charged. The multiple charges permit high molecular weight compounds to be observed with "low mass" instruments. There are three basic tandem mass analyzers used with ESI or NSI sources: (1) ion trap, (2) triple quadrupole (QQQ) and (3) quadrupole – time of flight (Q-TOF).

To elucidate even relatively simple cellular pathways, it is usually necessary to draw comparisons between several physiological conditions. Performing multiple pair-wise comparisons produces coverage gaps, since different peptides and proteins would be identified in each experiment. Therefore, a multiplexed quantitative technology not dependent on residue specific chemistry is advantageous toward differential profiling of large numbers of proteins. The introduction of an MS based isotopic tagging methods with cysteine-specific reagents such as ICAT Reagent or isobaric multiplex tagging methods with amine-specific stable isotope reagents such as iTRAQ Reagent (Applied Biosystems) enable quantitative proteomic analyses (Ross et al. 2004). In addition, an absolute quantification method using stable isotope labeling based on the work of Stemmann et al. (2001) was recently commercialized (AQUA, Thermo-Finningan). A multiplex protein absolute and relative quantification could be performed also directly (label-free) by Expression System (Waters) employing an advanced software (Protein Lynx Global Server, Waters).

19.4.3 MS Data for Protein Identification

For protein identification based on PMF, the measured masses of the cleaved peptides are compared with the calculated ones of theoretical digests of proteins contained in databases. As was mention previously, every peptidase used in proteomics research cleave protein in certain fashion. In the case of most widely used trypsin, it is behind the lysine and arginine. For this reason, every protein will yield to a specific number of tryptic peptides, based on number of lysines and arginines in the protein sequence. The computer is used to generate a list of theoretical tryptic peptides with their exact masses derived from each protein sequence available in the selected database and match them with the m/z values acquired by mass spectrometer. However, in situation where multiple proteins are isolated in one protein spot or for some other reason, such as low mass accuracy, or post-translational modifications, make PMF results questionable and in need of further validation or verification, MS/MS technique is applied. Upon obtaining MS/MS data from fragmenting a peptide two courses of action can be pursued. One is de novo sequencing, the other is search engine based identification. If the peptide is not contained in any database than de novo sequencing is a necessity. If the peptide however is expected to be present in a database it is more efficient to use a search engine. The reason is that complete de novo sequencing requires the present of all fragments corresponding to all cleavages. In reality this will only happen in circa 10% of all MS/MS experiments. Most

spectral data will be incomplete. The assumption that a peptide comes from a certain database reduces the solution space by many orders of magnitude and makes it possible to identify the true peptide sequence even with incomplete cleavage information. There exist a number of various search engines. The most widely used is SEQUEST (Yates et al. 1995a, b). It employs autocorrelation to estimate the similitude between the theoretical and experimental spectra. The overlap is expressed as a score which is the heart of all protein identification search engines. Mass spectrometry data derived from the unidentified protein are compared with theoretical data from known proteins, and a score is assigned according to how well the two sets of data matched. Any score above an arbitrary confidence threshold is termed a "hit." The top such hit is expected to identify the analysed protein. If there are no scores above this threshold ("no hits"), then the protein remains unidentified.

19.5 Conclusions

The main purpose of this chapter is to aid researchers in initiating proteomics experiments. Despite it short existence, proteomics already made a great impact in modern biology. Using PubMed search with keyword "proteomics", over 13,000 publications appears with oldest published only in 1998. In the case of plants, the number of proteome studies is rapidly expanding mainly due to completion of genomic sequences in organisms such as Arabidopsis thaliana. However, with development of MS methods, organisms with limited database resources are subject of proteomics investigations as well (Hajduch et al. 2006). Every proteomics experiment starts with protein isolation. Although numbers of protocols are available, optimal isolation procedure needs to be determined for each sample type individually. Prior protein identification, proteins in extracts must be separated to reduce complexity. Overall, proteome analysis is a powerful tool in functional characterization of plants. Proteomics approaches open up new perspectives for analysis of complex traits of crop species at various levels. In addition, proteomics data may be compared with transcriptomics data in order to understand the correlation between protein and transcript expression profiles. Further integration with transcriptomics will address important points in gene-protein relation and gene regulation in general, while integration with metabolomics will deliver new information related to regulation of metabolic processes on protein level. In near future, more tight correlation of proteomics data with other "omics" can be expected. These data will provide further in-depth proteome characterization of plant and subcellular compartments and potentially will identify new targets for crop improvements.

Acknowledgements The work was supported by FP7 of the European Union (MIRG-CT-2007-200165). This chapter reflects only the author's views and the Community is not liable for any use that might be made of information contained herein.

References

Adams P, Fowler R, Howell G, Kinsella N, Skipp P, Coote P, O'Connor CD (1999) Defining protease specificity with proteomics: a protease with a dibasic amino acid recognition motif is regulated by a two-component signal transduction system in Salmonella. Electrophoresis 20:2241–2247

Agrawal GK, Thelen JJ (2005) Development of a simplified, economical polyacrylamide gel staining protocol for phosphoproteins. Proteomics 5:4684–4688

Alban A, David SO, Bjorkesten L et al. (2003) A novel experimental design for comparative two-dimensional gel analysis, two-dimensional difference gel electrophoresis incorporating a pooled internal standard. Proteomics 3:36–44

Alexandrov ML,Gall LN, Krasnov NV, Nikolaev VI, Pavlenko VA, Shkurov VA (1984) Ion extraction from solutions at atmospheric pressure – a method for mass-spectrometric analysis for mass-spectrometric analysis of bioorganic substances. Dokl Akad Nauk SSSR 277:379–383

Berggren K, Chernokalskaya E, Steinberg TH et al. (2000) Background-free, high sensitivity staining of proteins in one- and two-dimensional sodium dodecyl sulfate-polyacrylamide gels using a luminescent ruthenium complex. Electrophoresis 21:2509–2521

Canas B, Pineiro C, Calvo E, Lopez-Ferrer D, Gallardo J (2007) Trends in sample preparation for classical and second generation proteomics. J Chromatogr A 1153:235–258

Chapman K (2002) The ProteinChip Biomarker System from Ciphergen Biosystems, a novel proteomics platform for rapid biomarker discovery and validation. Biochem Soc T 30:82–87

Chevallet M, Diemer H, Luche S et al. (2006) Improved mass spectrometry compatibility is afforded by ammoniacal silver staining. Proteomics 6:2350–2354

Choudhary G, Horvath C (1996) Ion-exchange chromatography. Methods Enzymol 270:47–82

Corbett JM, Dunn MJ, Posch A et al. (1994) Positional reproducibility of protein spots in two-dimensional polyacrylamide gel electrophoresis using immobilised pH gradient isoelectric focusing in the first dimension, An interlaboratory comparison. Electrophoresis 15:1205–1211

Davidssona P, Sjogrena M (2005) The use of proteomics in biomarker discovery in neurodegenerative diseases. Dis Markers 21:81–92

Delaplace P, van ver Wal F, Dierick J, Cordewener J, Fauconnier M, du Jardin P, America A (2006) Potato tuber proteomics: comparison of two complementary extraction methods designed for 2-DE of acidic proteins. Proteomics 6:6494–6497

Dolník V (2008) Capillary electrophoresis of proteins 2005–2007. Electrophoresis 29:143–156

Faurobert M, Mihr C, Bertin N, Pawlowski T, Negroni L, Sommerer N, Causse M (2007) Major proteome variations associated with cherry tomato pericarp development and ripening. Plant Physiol 143:1327–1346

Feng S, Ye M, Zhou H et al. (2007) Immobilized zirconium ion affinity chromatography for specific enrichment of phosphopeptides in phosphoproteome analysis. Mol Cell Proteomics 6:1656–1665

Isaacson T, Damasceno CMB, Saravanan RS, He Y, Catalá C, Saladié M, Rose JKC (2006) Sample extraction techniques for enhanced proteomic analysis of plant tissues. Nat Protocols 1:769–774

Görg A, Boguth G, Kopf A et al. (2002) Sample prefractionation with Sephadex isoelectric focusing prior to narrow pH range two-dimensional gels. Proteomics 2:1652–1657

Hajduch M, Ganapathy A, Stein JW, Thelen JJ (2005) A systematic proteomic study of seed filling in soybean. Establishment of high-resolution two-dimensional reference maps, expression profiles, and an interactive proteome database. Plant Physiol 137:1397–419

Hajduch M, Casteel JE, Hurrelmeyer KE, Song Z, Agrawal GK, Thelen JJ (2006) Proteomic analysis of seed filling in Brassica napus. Developmental characterization of metabolic isozymes using high-resolution two-dimensional gel electrophoresis. Plant Physiol 141:32–46

Haper RG, Workman SR, Schuetzner S et al. (2004) Low-molecular-weight human serum proteome using ultrafiltration, isoelectric focusing, and mass spectrometry. Electrophoresis 25:1299–1306

Issaq HJ, Conrads TP, Janini GM et al. (2002) Methods for fractionation, separation and profiling of proteins and peptides. Electrophoresis 23:3048–3061
Jensen ON, Wilm M, Shevchenko A, Mann M (1999) Sample preparation methods for mass spectrometric peptide mapping directly from 2-DE gels. Methods Mol Biol112:513–530
Jiang L, He L, Fountoulakis M (2004) Comparison of protein precipitation methods for sample preparation prior to proteomic analysis. J Chromatogr A 1023:317–320
Karas M, Bachmann D, Hillenkamp F (1985) Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. Anal Chem 57:2935–2939
Klose J (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. Humangenetik 26:231–243
Lamanda A, Zahn A, Roder D et al. (2004) Improved Ruthenium II tris (bathophenantroline disulfonate) staining and destaining protocolfor a better signal-to-background ratio and improved baseline resolution. Proteomics 4:599–608
Le Nel A, Minc N, Smadja C et al. (2008) Controlled proteolysis of normal and pathological prion protein in a microfluidic chip. Lab Chip 8:294–301
Lee WC, Lee KH (2004) Applications of affinity chromatography in proteomics. Anal Biochem 324:1–10
Lebowitz J, Lewis MS, Schuck P (2002) Modern analytical ultracentrifugation in protein science: a tutorial review. Protein Sci 11:2067–2079
Lescuyer P, Hochstrasser DF, Sanchez JC (2004) Comprehensive proteome analysis by chromatographic protein prefractionation. Electrophoresis 25:1125–1135
Madera M, Mechref Y, Novotny MV (2005) Combining lectin microcolumns with high-resolution separation techniques for enrichment of glycoproteins and glycopeptides. Anal Chem 77: 4081–4090
Manza LL, Stamer SL, Ham AJL et al. (2005) Sample preparation and digestion for proteomic analyses using spin filters. Proteomics 5:1742–1745
Mouradian S (2002) Lab-on-a-chip, applications in proteomics. Curr Opin Chem Biol 6:51–56
Natarajan S, Xu C, Caperna T, Garrett W (2005) Comparison of protein solubilization methods suitable for proteomic analysis of soybean seed proteins. Analytical Biochem 342:214–220
Neuhoff V, Arold N, Taube D et al. (1988) Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. Electrophoresis 9:255–262
Nilsson CL, Davidsson P (2000) New separation tools for comprehensive studies of protein expression by mass spectrometry. Mass Spectrom Rev 19:390–397
O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. J Biol Chem 250:4007–4021
Peng Y, Pallandre A, Tran NT et al. (2008) Recent innovations in protein separation on microchips by electrophoretic methods. Electrophoresis 29:157–178
Pons L, Chéry C, Mrabet N et al. (2005) Purification and cloning of two high molecular mass isoforms of peanut seed oleosin encoded by cDNAs of equal sizes. Plant Physiol Biochem 43:659–668
Posewitz MC, Tempst P (1999) Immobilized gallium (III) affinity chromatography of phosphopeptides. Anal Chem 71:2883–2892
Rabilloud T (2002) Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. Proteomics 2:3–10
Rabilloud T, Strub JM, Luche S et al. (2001) A comparison between Sypro Ruby and ruthenium II tris (bathophenanthroline disulfonate) as fluorescent stain for protein detection in gels. Proteomics 1:699–704
Rathinasabapathi B, Fouad WM, Sigua CA (2001) β-Alanine betaine synthesis in the plumbaginaceae. purification and characterization of a trifunctional, S-adenosyl-l-methionine-dependent N-methyltransferase from limonium latifolium leaves. Plant Physiol 126:1241–1249

Rehm H (2006) Protein biochemistry and proteomics. Academic, London
Righetti PG, Castagna A, Antonioli P et al. (2005) Prefractionation techniques in proteome analysis, The mining tools of the third millennium. Electrophoresis 26:297–319
Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, Pappin DJ (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 12:1154–1169
Shevchenko A, Wilm M, Vorm O et al. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem 68:850–858
Song J, Braun G, Bevis E, Doncaster K (2006) A simple protocol for protein extraction of recalcitrant fruit tissues suitable for 2-DE and MS analysis. Electrophoresis 27:3144–3151
Sriyam S, Sinchaikul S, Tantipaiboonwong P et al. (2007) Enhanced detectability in proteome studies. J Chromatogr B 849:91–104
Stasyk T, Huber LA (2004) Zooming in, Fractionation strategies in proteomics. Proteomics 4:3704–3716
Steinberg TH, Agnew BJ, Gee KR et al. (2003) Global quantitative phosphoprotein analysis using Multiplexed Proteomics technology. Proteomics 3:1128–1144
Stemmann O, Zou H, Gerber SA, Gygi SP, Kirschner MW (2001) Dual inhibition of sister chromatid separation at metaphase. Cell 107:715–726
Striegel A (2008) Size-exclusion chromatography, smaller, faster, multi-detection, and multi-dimensions. Anal Bioanal Chem 390:303–305
Tastet C, Lescuyer P, Diemer H et al. (2003)A versatile electrophoresis system for the analysis of high- and low-molecular-weight proteins. Electrophoresis 24:1787–1794
Tirumalai RS, Chan KC, Prieto DA et al. (2003) Characterization of the low molecular weight human serum proteome. Mol Cell Proteomics 2:1096–1103
Tishchenko G, Dybal J, Mészárosová K et al. (2002) Purification of the specific immunoglobulin G1 by immobilized metal ion affinity chromatography using nickel complexes of chelating porous and nonporous polymeric sorbents based on poly(methacrylic esters) effect of polymer structure. J Chromatogr A 954:115–126
Zellner M, Winkler W, Hayden H et al. (2005) Quantitative validation of different protein precipitation methods in proteome analysis of blood platelets. Electrophoresis 26:2481–2490
Zhang H, Yan W, Aebersold R (2004) Chemical probes and tandem mass spectrometry, a strategy for the quantitative analysis of proteomes and subproteomes. Curr Opin Chem Biol 8:66–75
Yuana X, Desiderioa DM (2005) Proteomics analysis of human cerebrospinal fluid. J Chromatogr B 815:179–189
Vorderwülbecke S, Cleverley S, Weinberger SR et al. (2005) Protein quantification by the SELDI-TOF-MS–based ProteinChip System. Nat Methods 2:393–395
Wang S, Zhang X, Regnier FE (2002) Quantitative proteomics strategy involving the selection of peptides containing both cysteine and histidine from tryptic digests of cell lysates. J Chromatogr A 949:153–162
Wang H, Hanash S (2003) Multi-dimensional liquid phase based separations in proteomics. J Chromatogr B - Analyt Technol Biomed Life Sci 787:11–18
Wang W, Vignani R, Scali M, Cresti M (2006) A universal and rapid protocol for protein extraction from recalcitrant plant tissues for proteomic analysis. Electrophoresis 27:2782–2786
Wehr T (2001) Separation Technology in Proteomics. LCGC 19:702–711
Wilkins MR, Sanchez J-C, Gooley AA, Appel RD, Humphrey-Smith I, Hochstrasser DF, Williams KL (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotechnol Genet Eng Rev 13:19–50
Wolters DA,Washburn MP, Yates JR (2001) An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem 73:5683–5690
Wu S, Karger B (1996) Hydrophobic interaction chromatography of proteins. Methods Enzymol 270:27–47

Wu J, Lenchik NJ, Pabst MJ et al. (2005) Functional characterization of two-dimensional gel-separated proteins using sequential staining. Electrophoresis 26:225–237

Yates JR 3rd, Eng JK, McCormack AL (1995a) Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. Anal Chem. 67:3202–3210

Yates JR 3rd, Eng JK, McCormack AL, Schieltz D (1995b) Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. Anal Chem 67:1426–1436

Chapter 20
Metabolomics: Novel Tool for Studying Complex Biological Systems

Federica Maltese and Robert Verpoorte

Abstract The new era of functional genomics is leading towards the discovery of gene functions. Genomics, transcriptomics, proteomics and metabolomics aim to acquire a comprehensive and integrated understanding of living organisms. Metabolomics defines the phenotype of an organism from the chemical point of view, by analysing the ultimate products of the entire metabolic network at a certain point in time. Plant metabolomics is a powerful tool in systems biology and functional genomics studies, it has already been applied for agricultural purposes (screening of progeny stability, analysis of plant-pathogen interaction), in substantial equivalence studies of GMO and in quality evaluation of food, beverages and phytomedicines. The range of metabolites in an organism is so diverse that so far none of the analytical technologies applied for metabolomics studies purpose result to be adequate, and a combination of them must be used

20.1 Introduction

Dixon et al. (2006) wrote: "biological systems are exceedingly complex". To understand these systems is a major challenge for mankind.

With the achievement of the complete genome sequence of bacteria first, then humans, and now green plants as well (*Arabidopsis* and rice), science has made a first step in the understanding of the function of the entire gene set of living systems: the era of functional genomics has started. Although the complete sequence of some organisms has been obtained, still 20–40% of the open reading frames cannot be annotated (Goodacre 2005). In the case of the *Arabidopsis* genome, only 9% of the genes have been characterized (Haas et al. 2005). This has led to a renewed interest into the correlation between genes and their phenotypic expression. The traditional dogma of molecular biology (from gene to transcript to protein) has been put aside with the new concept of systems biology. The flow of information deriving from

F. Maltese and R. Verpoorte (✉)
Department of Pharmacognosy, Section of Metabolomics, Institute of Biology, Leiden University, Leiden, The Netherlands
e-mail: verpoort@lacdr.leidenuniv.nl

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_20,

genes is in reality networked by many loops with their downstream products, resulting in a complex and dynamic system of transcripts, proteins and metabolites. As a consequence, changes in the quantities of individual enzymes might have little effect on metabolic fluxes through pathways, but may have significant effects on the concentrations of individual metabolites (Goodacre et al. 2004).

Genomics, transcriptomics, proteomics, and metabolomics, are the new platforms for the development of functional genomics. The goal of the "omics" approaches is to acquire a comprehensive and integrated understanding of biology. Combined information from these technologies is needed to help us in understanding a cell at the system level.

Metabolites are the downstream products of gene expression. The analysis of the metabolome of an organism is expected to provide the most direct link between genome and phenotype, thus dealing with the most important question of functional genomics. In this perspective, metabolomics has been suggested to provide the most "functional" information among the "omics" technologies (Sumner et al. 2003).

20.2 Characteristics of Metabolomics

Taking a look at the quantitative dimensions of the different "omics" elements in the context of human genome (Fig. 20.1), one can perceive that measuring the metabolome of an organism is the most straightforward way to understand the complexity of the biological system.

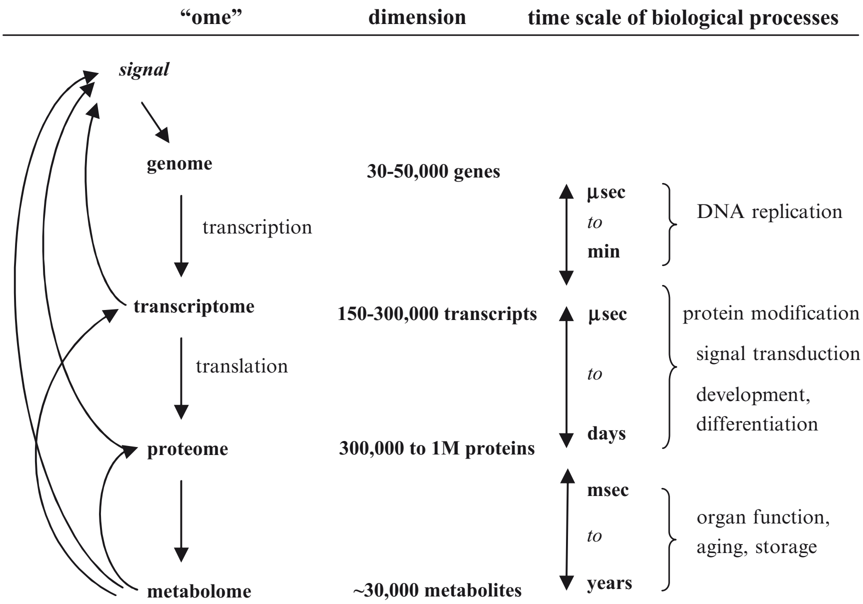


Fig. 20.1 Complexity of the different "omics" levels with respect to quantitative dimensions and time scales of processes within an organism. Adapted from Lay et al. (2006)

Metabolomics is the comprehensive identification and quantification of all low-molecular weight molecules (<3,000 *m/z*) present in a cell or tissue under a given set of conditions (Goodacre et al. 2004). Different metabolomics-related strategies are currently employed, none of which fulfills the exact objectives of metabolomics. These are:

- Metabolite profiling, which is focused on a set of pre-defined metabolites, generally associated with a specific pathway
- Metabolic fingerprinting, a radid high-throughput analysis which is used to classify samples (normally based on their provenance or origin)
- Metabolite targeted analysis, which is restricted to one or two metabolites related to specific metabolic reaction that would be affected by biotic or abiotic perturbation

A term that can lead to confusion is metabonomics, which is used for NMR-clinical applications on tissues or biological fluids whose change in metabolite levels is the result of disease or therapeutic treatment.

After the term "metabolome" was first introduced to describe the set of metabolites synthesized by an organism (Oliver et al. 1998), and "metabolomics" was for the first time defined (Fiehn 2001), the literature in this field has showed an exponential growth. Metabolomics is nowadays applied to a wide variety of disciplines. In the field of functional genomics, we have already discussed the special role played by metabolomics in deciphering the gene function by establishing a better understanding of the correlation between genes and the functional phenotype of an organism (Bino et al. 2004).

In the context of systems biology, the potential of the metabolomics platform has just started to be exploited. The general system biology approach consists of a perturbation of the system (biologically, genetically, or chemically), followed by monitoring the impact of the perturbation at the genomics, proteomics, and metabolomics levels (Ryan and Robards 2006). Metabolomics is the core of system biology, as it is best suited for the systems biology approach owing to its comprehensive information content concerning dynamic metabolic networks (Weckwerth 2003). By many criteria, the concentration of different metabolites in tissues or body fluids is closely linked to the variations in phenotype that are most relevant to human growth, development and health (Watkins and German 2002). Thus, global "omics" approaches to metabolite analyses are likely to be the most valuable tools for studying various approaches to improve human health through diet or drugs. In pharmacology, the metabonomics has already started to be applied for toxicology, drug discovery and human health assessment. The concept of biomarkers has been introduced: the metabolome fingerprinting of mammalian disease models versus control should be used to discover early biomarkers of disease. Application of biomarkers discovery is at its infancy for disease prognoses, diagnoses, and therapy monitoring, but looks very promising (Goodacre 2007).

As in the analysis of the effects of toxicological and pharmaceuticals, metabolomics can be used for the evaluation of the effects of nutritionals on the human metabolome. The emerging science called "nutrigenomics" studies how specific genes and bioactive food component interact, with the aim of "prescribing" individualized diets using "functional foods" (Afman and Muller 2006; Mariman 2006;

Gibney et al. 2005). Obviously understanding the effects of food on health also requires knowledge of all metabolites in our food, i.e. plant metabolomics will play a major role in this field.

20.3 Plant Metabolomics

The aim of plant metabolomics is to analyse all the products of the entire metabolic network at a certain point in time. The size of the total metabolome in the plant kingdom has been estimated to be about 200,000 primary and secondary metabolites (Fiehn 2002). About 150,000 have been identified (Encyclopedia of Natural Products) and ca. 4,000 new are reported every year (Verpoorte et al. 1999). But in fact the number might be much bigger, assuming that the 250,000 species all have specific secondary metabolite pathways. The size of the metabolome of a plant is also difficult to estimate. If we assume that this is in the same order as the number of genes, it would be ca. 30,000 compounds. This means a great challenge for the scientists to accomplish the main aim of metabolomics: to take a snapshot of the metabolome qualitatively and quantitatively.

Despite the many sophisticated analytical technologies that can be used, so far, none of the methods proposed for the profiling of all the metabolites in a plant (or other organisms) proves to be adequate. Plant metabolites are too many, chemically too complex and too much different in chemical characteristics and occur in a very broad dynamic range of concentrations. All this implies that the analytical methods must be selective for the broadest possible range of metabolites, and sensitive to fluctuation in their quantity. As a result, no single technique is suitable for the analysis of the complete plant metabolome, but a combination of techniques must be used (Weckwerth 2003).

Another characteristic of metabolomics is the unbiased approach. "Omics" technologies aim to the non-targeted identification of all gene products produced directly or indirectly (i.e. transcripts, proteins and metabolites) and present in a specific biological sample (Fridman and Pichersky 2005).

The required unbiased nature of metabolomics analysis is already challenged at the initial step of the analytical workflow: the sampling. All the information about the living organism must be taken in consideration and, when possible, controlled before starting the experiment in order to reduce experimental errors. The sampling method can influence the reproducibility of the analytical sample (Dunn and Ellis 2005; Verpoorte et al. 2008). Bino et al. (2004) proposed a comprehensive protocol in the MIAMET (Minimum Information About a Metabolomics Experiment), which includes the important steps of the experimental design. After the problem of harvesting the plant material, follows the extraction process. Polar/non-polar extractions are the most frequently applied method and are performed by physical/chemical disruption of the cells, removal of the cell pellet by centrifugation and distribution of metabolites to polar (methanol/water) and non-polar (chloroform) solvents (Dunn and Ellis 2005). In any case, there is no extraction method which

does not imply an intrinsic bias towards certain groups of metabolites. Once again, the unbiased nature of metabolomics is violated by choosing a certain extraction procedure. The same consideration is true for the subsequent detection technique employed. For a comprehensive description of the critical issues relating to the generation of metabolomics data, we refer to other authors (Ryan and Robards 2006a; Kim et al. 2006; Verpoorte et al. 2008). Here we will briefly illustrate some characteristics and limitations of the most common metabolomics tools.

20.4 Analytical Tools

20.4.1 Methodologies Employed

Comprehensiveness, sensitivity, selectivity, reproducibility, and short time of analysis are the required parameters for the analytical tools employed in metabolomics. The technology platforms used in plant metabolomics can be categorized in chromatographic or spectrometric methods. GC, LC, capillary electrophoresis (CE) are included in the chromatographic methods, whereas MS and NMR belong to the other group. As a general consideration the chromatographic techniques require some preliminary steps before the measurement, such as pre-fractionation or derivatization. Consequently, the spectrum of metabolites analyzed is not as comprehensive as it is for spectrometric methods. Also reproducibility is a problem connected to chromatographic separation methods, different column types from different manufacturers do not give the same result, and the characteristics of a certain type of column will change over time of production, as well as of use. In most cases, the above mentioned chromatographic methods are coupled to MS. GC-MS and LC-MS are the most popular analytical tools in metabolomics, while CE-MS is gaining importance. The strategy is to separate the metabolites before analyzing them in the mass spectrometer.

In GC-MS the high sensitivity of MS (at "pg" level) is coupled to the high separation efficiency of GC. But only volatile compounds can be submitted to GC analysis. A derivatization step prior to the analysis, like sililation, is required for hydrophilic or thermolabile metabolites. The generated chromatograms become much more complicated by the many derivatives yielded by metabolites with more than one reactive site. To overcome this problem, the introduction of fast acquisition rate time-of-flight (TOF) instruments, sometime coupled with deconvolution softwares, optimizes the identification of all the compounds in the chromatogram (Dunn and Ellis 2005). High mass resolution, together with high mass accuracy, is guaranteed by the employment of Fourier-transform ion cyclotron resonance, (FT-ICR)-MS, which is now applied to the metabolomics field (Brown et al. 2005; Murch et al. 2004; Tohge et al. 2005). The GC-MS method is supported by commercially available databases for compound identification and spectra elucidation. Although not covering the huge number of natural metabolites, those databases represent user-friendly tools for structural confirmation of metabolites, particularly for the well known primary metabolites.

LC-MS is considered an important additional technology for plant metabolomics (Hall 2006). Molecules which are not detectable by GC, either because they cannot be made volatile or because they are unstable during the derivatization, or at the high temperatures in the GC, can be detected by LC. Aqueous samples can be analyzed by LC, with minimal sample preparation or even directly. The LC-MS analysis requires that the compound can be ionized, either as a positively or negatively charged ion. This can be achieved by electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), atmospheric pressure ionization (API), and atmospheric pressure photoionization (APPI). New ESI-MS instruments operate switching continuously between positive and negative modes during a run. This allows that the broadest range of molecules can be detected. To give more robust results, LC can be combined with tandem mass spectrometry (LC/MS/MS). The MS-fragmentation pattern observed add to the selectivity of the system, and thus to the correct identification of the peaks. A major limitation in using LC-MS is that the technology is not yet supported by an extensive spectral database, thus the metabolite identification step is more time consuming than in GC-MS. Furthermore, among the chromatographic methods, LC has the smallest peak capacity, meaning a reduced resolution of individual metabolites in the mixture (Fukusaki and Kobayashi 2005). Nevertheless, the last advances in column performance have improved LC application in metabolomics. Another strategy that has been introduced to overcome the peak resolution limitation is the introduction of a second dimension. Comprehensive 2-dimensional LC chromatography (LC x LC) is normally employed in proteomics studies, but it has demonstrated to be a powerful technique also when applied to plant metabolomics research (Mondello et al. 2005; Dugo et al. 2006; van der Klift et al. 2008). The separation of the entire sample is highly improved, since all the fractions eluted in the first dimension are submitted to a second separation. Some other technical problems arise in regards to mobile phase immiscibility, precipitation of buffer salts and incompatibility of stationary phase (Ryan and Robards 2006b). Nevertheless, LC × LC is very promising in metabolomics studies. In an analogues way as comprehensive 2-dimensional LC, GC has also been developed as GC × GC for the characterization of complex samples, but with very few examples for plant extract analysis (Shellie et al. 2003; Hope et al. 2005; Pierce et al. 2006).

Capillary electrophoresis (CE) is gaining popularity due to the ease of application, the reduced costs, and the small volumes required (less than nL). It is based on the different electrophoretic mobilities of ions in electrophoretic media. Capillary zone electrophoresis (CZE) and micellar electrokinetic chromatography (MEKC) have been applied for the metabolomic analysis of small molecules, such as microbial extracts (Markuszewski et al. 2003; Soga et al. 2003; Panagiotou et al. 2005). CE offers high-resolution chromatographic separation coupled to sensitive detection when combined with MS. For this reason we can expect it will be more widely applied to the analysis of plant extracts, being suitable for the detection of a diverse range of primary and secondary metabolites (Sato et al. 2004).

Yet, all the chromatographic based methods have intrinsic biases against certain classes of compounds (Weckwerth 2003). The employment of combinations of two or more separation methods with different selectivity and the application in parallel

to the same sample is a strategy to solve this limitation (Sumner 2006). Another major limitation is represented by the need of construction of calibration curves, as all compounds will leave a different detector response. Consequently absolute quantification of all metabolites can not be achieved by chromatography, only for each single compound percentage of changes can be determined.

Nuclear magnetic resonance (NMR) spectroscopy is presently the method of choice for medical metabolomics (metabonomics). Changes in the metabolites composition of biofluids are directly investigated as resulting from disease conditions, drug activity or drug toxicity. Important biomarkers can thus be identified for the comparison of the healthy vs diseased humans. Despite the lesser sensitivity in comparison to GC and LC methods, NMR is widely applied to the analysis of plant metabolites. It is rapid, non-destructive, reproducible, and it requires very minimal sample preparation (Lindon et al. 2003). Spectra can be recorded from crude extracts without derivatization, making the composition of the sample closer to the actual levels in tissues (Last et al. 2007). NMR also allows quantitative assessment of single metabolites present in the mixture, and it represents a direct structure elucidation tool as well. Unlike chromatographic or mass spectrometry methods, NMR is not selective for certain class of analyzed compounds. The requisite is for the molecules to bear a "magnetically active" nucleus, normally ^{1}H, because of its high natural abundance, but it can also be ^{13}C, ^{15}N, or ^{31}P. Which means that every kind of organic molecule can be detected using NMR. Furthermore, the introduction of NMR detectors cooled to near-cryogenic temperatures (cryoprobes) make it possible to gain much higher sensitivity, and this to measure smaller amounts of sample or to use less time (Lindon et al. 2003).

^{1}H NMR spectra contain several hundred signals which are limited to a very small chemical shift range (about 10 ppm). Furthermore, the instrument is relatively not sensitive, so that signals from compounds present at low concentrations may not be detected. As a result, the one-dimensional (1D) spectra show many overlapping peaks which hamper the identification of the compounds (Fig. 20.2).

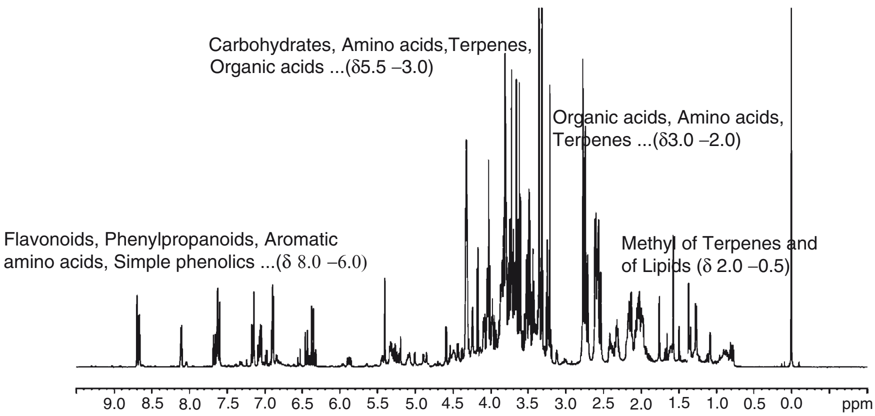


Fig. 20.2 Complex ^{1}H NMR spectrum of tobacco leaf extracted by MeOD:KH_2PO_4 buffer in D_2O (1:1), pH 6.0 (500 MHz)

Two-dimensional (2D) NMR experiments represent, however, a major tool for the resolution of overlapping signals, and for the characterization of unidentified compounds following to a first metabolic fingerprinting obtained with the traditional 1D NMR. Among other 2D techniques, J-resolved spectra require relatively short time of measurement, and show interesting applications in resolving complex spectra of mixtures, allowing the identification of minor compounds. In this kind of spectra the second dimension gives for each signal the coupling constant. J-resolved represents a useful deconvolution tool in NMR-based metabolomics (see below). The result is a less complex spectrum, while the resolution is higher (Fig. 20.3 and 20.4) (Liang et al. 2006; Hendrawati et al. 2006).

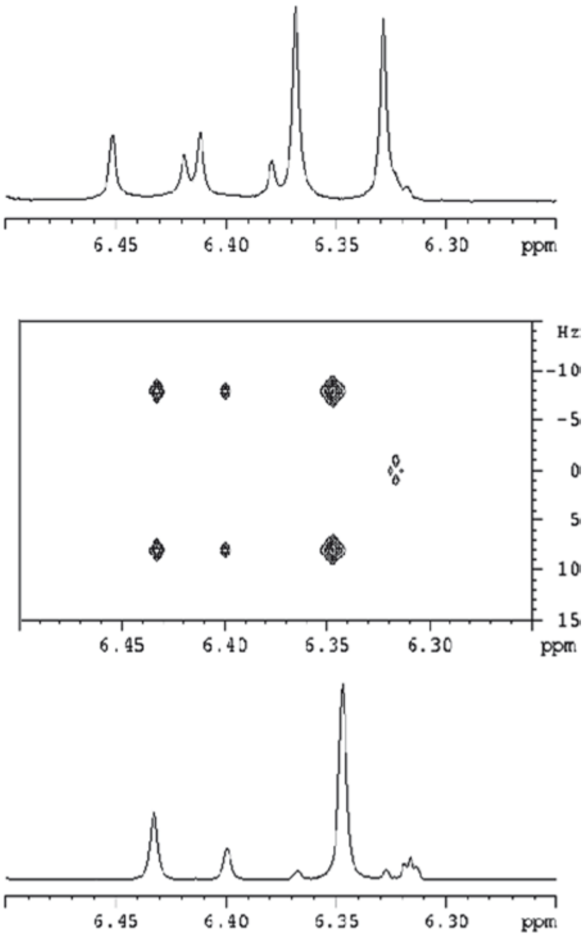


Fig. 20.3 J-resolved spectra in the region between δ 6.25 and δ 6.50 of tobacco leaf extracted by MeOD:KH_2PO_4 buffer in D_2O (1:1), pH 6.0 (500 MHz). In the upper trace the original ^1H NMR is shown. In the middle trace the 2D J-resolved spectrum reveals the presence of three doublets (overlapping in the original ^1H NMR spectrum). In the lower trace the original signals are projected in 1D as singlets

Another possible application of NMR to metabolomics studies is the stable isotope (^{13}C, ^{31}P or ^{15}N) labeling. This technique can be applied to enhance the signal of low natural abundant isotopes, and it can thus provide a powerful tool for the metabolic flux analysis (Kruger et al. 2003; Ratcliffe and Schachar-Hill 2005; Mustafa 2007). This is also very useful in unraveling biosynthetic pathways, through a retrobiosynthetic approach (Werner et al. 1997).

20.4.2 Data Processing and Mining

The complexity of the spectra obtained either with NMR or MS requires elaboration of the data using chemometrics tools. The visual inspection of those spectra can in fact only give a very minimal view of all the information included. The application of multivariate statistical techniques (chemometrics) is thus required to obtain the maximum of information from such complex data. The raw data produced as chromatograms or NMR spectra must be first converted in a "clean" format which is suitable for the analysis. This step is called data processing, and it represents one of the most crucial and time-consuming steps of the entire data analysis. Signal deconvolution, phasing, baseline correction, noise suppression, peak alignment, peak picking and quantification, they all guarantee a reliable comparative metabolomic analysis based on the sample-to-sample variability loss. In this way the raw instrumental data are cleaned up and ready to be submitted to the multivariate analysis (MVA).

A chromatogram, or a NMR spectrum are characterized by several hundreds peaks. These all represent variables in the data matrix, which is also composed by the observations (samples). Metabolomics data must be considered as multivariate in the extent that they consist of these matrixes. Like other "omics" data sets, they contain many more variables than samples. Sometimes the number of variables can be reduced by applying a bucketing procedure. The whole data set space is divided in equally sized subset of data (*buckets* or *bins*) and the intensities are summed within each of them (Holmes et al. 1994). For example, in the case of NMR data, the number of variables can be reduced to about 250 bins of 0.04 ppm width. The bucketing procedure of NMR data can also avoid misinterpretations caused by peak shifts due to pH effect.

The aim of metabolomics analysis is to classify samples into different groups (e.g. genotypes or treatments) or predict its origin. In this context the MVA can be applied, in the extent that it extracts the information from data containing multiple variables, by using all the variables simultaneously. MVA can be divided into two categories: the "unsupervised" and the "supervised" methods. The unsupervised multivariate techniques are used to establish any existing clustering within the data set. The most commonly used unsupervised methods are principal component analysis (PCA) and hierarchical cluster analysis (HCA), but others are also routinely applied, such as non-linear mapping (NLM), and a self-organizing map (SOM). PCA can reduce the number of variables needed to describe the variance in the data set. Mathematically based on eigenvectors, the PCA generates a set of lines and planes that closely describe the system of the original matrix of data formed by N rows

(the observations) and *K* columns (the variables). The first principal component (PC1) is the line in the *K*-dimensional space that approximates the best the data in the least squares sense (Eriksson et al. 2006). In other words, PC1 defines the direction of maximum spread (variance) in the data set. Then the second component (PC2) is calculated. This is orthogonal to the first one, and improves the approximation of the *X*-data, defining the direction of greatest remaining spread (Colquhoun 2007). The two components originate a plane in which the projections of all the observations are identified by co-ordinates (*scores*). The score plot represents the graphical visualization of the plane, and shows similarities between the observations. The model can calculate as many PCs as necessary to summarize the information in a data set. The loading plot is geometrically built on the orientation of the PCs in relation to the original variables. The loading plot allows the evaluation of the influence of each variable (metabolite) onto the PCs (which variables are responsible for the pattern? How these variables ARE correlated to each other?), and to the total information of the entire metabolome. PCA may reveal groups of observations, trends, or outliers, also showing the relationship between observations and variables (Eriksson et al. 2005).

HCA is a clustering method in which the distance between each pair of samples is calculated and the pair that is most similar is identified. The midpoint between this pair is calculated and the distance between it and all the subsequent samples included in the analysis is calculated. The measurement can be repeated all the samples are included in a data set. The result is visualized as a tree-like diagram called dendrogram (Lindon et al. 2003).

Supervised classification methods are applied when the classes of the response that one is trying to predict for each of the samples are known. Then the aim is to identify a model that will correctly associate all the inputs with the targeted class (Goodacre et al. 2007). Partial least squares (PLS), linear discriminant analysis (LDA), soft independent modeling of class analogy (SIMCA), k-nearest neighbour analysis (KNN), and artificial neural networks (ANNs) are supervised methods generally applied to the analysis of metabolomics data. PLS is successfully used when the number of variables is much greater than the number of samples. In an analogous way as in the PCA model, PLS is also constituted by lines and planes. In this case the line orientations are calculated in order to optimally discriminate between groups (Colquhoun 2007). The *X*-data-matrix formed by variables and observations is correlated with a *Y*-matrix, artificially created to describe the class memberships of the observations. For each matrix, *X* and *Y*, the calculation of PLS-components defines planes. The observations are represented by one point in the *X*-matrix and one point in the *Y*-matrix. This allows to identify any correlation of the points in the two matrixes.

SIMCA classifies and predicts unknown samples in a PCA based way, so that any sample is tested for being included or not in a specific group or category. SIMCA builds a PC model for each class, and when a sample does not satisfy the model, a new PC is constructed and the model validated.

For detailed discussion of these and other bioinformatics tools is referred to Lindon et al. 2001; Eriksson et al. 2006; Berrueta et al. 2007.

20.5 Applications of Plant Metabolomics

Plant metabolomics has a wide range of possible applications.

As one of the complementary analytical method in functional genomics, the metabolomics based platform is an unsurpassed tool in the hands of plant physiologists and plant breeders. Based on the assumption that any alteration in the gene regulation will be reflected in alterations of biochemical pathways, the metabolomics analysis will help in the elucidation of gene functions, signalling and metabolic networks. In this sense, the consequences of certain treatments (stress or perturbation), genetic variation, or environmental changes, can be identified through phenotypic differences. An example is the work of Cook et al. (2004) in which GC-TOF-MS is applied to measure the extent of the role of *Arabidopsis* CBF cold response pathway in terms of metabolome changes when the plant is stressed by low-temperature conditions. In another example, GC-TOF-MS based metabolomics was applied for the analysis of silent potato mutants, showing the potential of this strategy to distinguish the silent phenotype from the parental line (Weckwerth et al. 2004). Nutritional stress on crops has also been investigated using different transcriptomics – metabolomics approaches (Nikiforova et al. 2004; Hirai et al. 2004). For a review of metabolomics methods applied to study plant stress responses, see Shulaev et al. (2008).

What metabolic fingerprinting also offers is mapping of natural genetic variations in crops (Keurentjes et al. 2006) or rapidly screening the progeny stability of some crops. Both strategies can be used as tools for selection of more resistant progenies, essential for breeders (Hall 2006). Very important from the commercial point of view of crops is also the threat represented by parasites in field. The metabolic changes in indolic compounds and phenylpropanoids of wild type and mutant root cultures of *Arabidopsis* infected by the root-pathogenic oomycete *Pythium sylvaticum* were determined in an extensive work based on ^{1}H-NMR, ESI-MS, APCI-MS coupled with quadrupole time of flight (QTOF)-MS (Bednarek et al. 2005).

The effects of phytoplasma infection on metabolomic changes in *Catharanthus roseus* leaves were investigated using 1D and 2D NMR spectroscopy, showing that the infected leaves were characterized by increased levels of terpenoid indole alkaloids, phenylpropanoids, and several primary metabolites. The metabolic pathways linked to these metabolites were considered by the authors as involved in the defence response of the plant (Choi et al. 2004a). The same authors applied the methodology for the metabolic discrimination of tobacco plants infected by tobacco mosaic virus (TMV) in order to elucidate the metabolic changes connected to systemic acquired resistance (SAR) (Choi et al. 2006). A complex picture of the metabolic alterations caused by virus infection was obtained, also considering different time points, and different developmental stages of the leaves. The effect of the interaction of different bacteria strains on the metabolome of *Brassica rapa* was also investigated by mean of NMR spectroscopy by Jahangir et al. (2007). The results demonstrated the potential of NMR-based metabolomics as a tool to study the interaction of food-borne bacteria and vegetables.

The rapid and comprehensive analysis of a plant metabolome has also developed in an important tool in the screening of genetically modified organisms (GMO).

The safety evaluation of GMO is gaining importance with the current international policies which require the certification of new crops before releasing the license for their commercialization. In particular this is important in food safety assessments. A large scale metabolomic analysis of field grown GM potato plants has been published by Catchpole et al. (2005). A multi-targeted approach (LC-MS, GC-TOF-MS) has been applied for the metabolome fingerprinting of field-grown tubers from conventional potato cultivars and genotypes bioengineered to contain high levels of inulin-type fructans. Conclusions were drawn with respect to the concept of "substantial equivalence" which is used as starting point to assess the safety of genetically modified food (Kuiper et al. 2002). The metabolomic analysis is extremely useful in substantial equivalence studies. The safety of a GM crop is assessed on the base of its change in the concentration of any metabolite. If the change is within the natural range of concentrations found in natural cultivars, the change would be regarded as "safe" (Colquhoun et al. 2006). NMR has thus revealed to be a powerful tool for substantial equivalence studies of transgenic food crops in comparison with traditionally bred cultivars (Noteborn et al. 2000; Le Gall et al. 2003; Charlton et al. 2004; Manetti et al. 2004).

Linked to the evaluation of GMO, there are two other possible applications of metabolomics: one in optimizing metabolic engineering techniques, the other one in the study of the unintended effects of genetic manipulation. Considering the assumption that gene-proteins-metabolites are not unidirectionally linked, but instead they are networking through many pathways and feedbacks, it is clear that manipulating one metabolic pathway may lead to multiple and unexpected consequences. The comprehensive metabolic profiling of the effects of genetic manipulations can lead to understanding the whole system and to discover key control steps in a target pathway (Harrigan et al. 2007).

A comprehensive analysis of some *Arabidopsis* mutants containing an antisense construct for the *chalcone synthase* gene was performed using a combination of LC-MS, LC-DAD, and NMR (Le Gall et al. 2005). The effects of the genetic modification on the levels of flavonoids, sinapate esters, and glucosinolates were observed, and the interfering effects of genetic and environmental variation were also considered.

Metabolomics has a huge potential in product fingerprinting, particularly for quality evaluation of food and beverages, and detection of adulterations. In this field the technology, especially NMR based, is already widely applied (Vogels et al. 1996; Le Gall et al. 2001; Brescia et al. 2002; Charlton et al. 2002).

In the same way, quality control of phytomedicines has started to include metabolomics approaches. The composition of chamomile flower extracts of three different geographic origins has been studied by Wang et al. (2004). In this case ^{1}H NMR spectral analysis was also applied to determine the amount of stalk material which was added to the chamomile flowers. The metabolic fingerprints of willow bark extracts from different origin has been characterized (Hendriks et al. 2005), and the differences in metabolic composition of St. John's wort extracts have been investigated by ^{1}H NMR and correlated with the activity on the opioid receptor (Roos et al. 2004). An NMR based study was also used by Frederich et al. (2003)

to quantify strychnine and brucine levels in different parts of various *Strychnos* species. A similar strategy was adopted to discriminate twelve cultivars of *Cannabis sativa* (Choi et al. 2004b). Kim et al. (2005) evaluated the applicability of NMR based metabolic fingerprinting to the analysis of three different *Ephedra* species and nine commercial *Ephedra* herbs. The method could be used for chemotaxonomic analysis and quality control purposes.

Furthermore, it is nowadays clear that the reductionist approach for the purpose of activity studies on traditional medicine (i.e. traditional Chinese medicine – TCM- and Ayurveda) does not match the holistic approach of Chinese medicine (Verpoorte et al. 2005). It is known that phytopharmaceuticals, and TCM in particular, are often mixtures of different plants, mainly working in synergism, and that their composition is difficult to control. Systems biology offers the possibility to understand the mode of action of phytomedicines by comparing the changes in the transcriptome, proteome and metabolome patterns with those observed with known drugs. In this context, metabolomics can link the response of living organisms to the activity of a compound or a combination of compounds, thus possibly identifying new modes of action, pro-drugs and synergisms in the complex mixtures of compounds in traditional medicine (Wang et al. 2004; Verpoorte et al. 2005). Recently, the metabolomic profile of a Mexican medicinal plant (*Galphimia glauca*) was determined by using NMR and multivariate data analysis was applied for differentiating populations growing in different geographical areas (Cardoso-Taketa et al. 2008). The results were then correlated with those of two neuropharmacological assays, demonstrating the efficacy of using metabolomics for the in silico identification of active principles in medicinal plants.

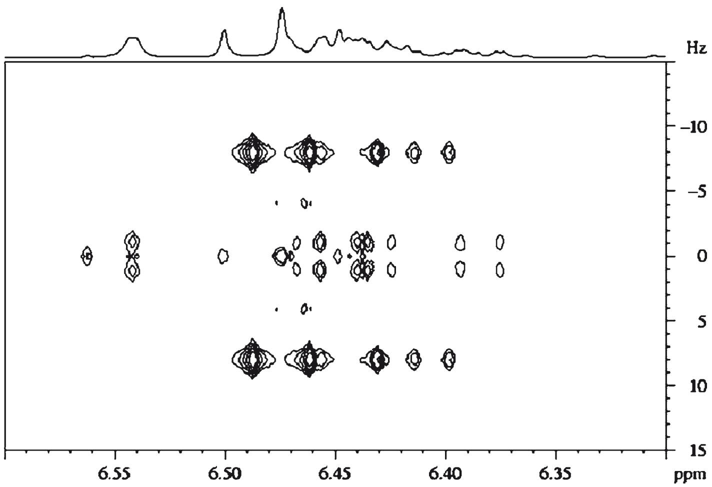


Fig. 20.4 An example of J-resolved spectrum in the region between δ 6.30 and δ 6.60 of *Brassica rapa* extracted by MeOD:KH_2PO_4 buffer in D_2O (1:1), pH 6.0 (600 MHz). Low and overlapping signals in the aromatic region of the spectrum are resolved in the second dimension

20.6 Conclusions

Metabolomics, as it has been defined, is a very challenging field. This platform is demonstrating to be a powerful tool in several applications, such as functional genomics, systems biology, biomarker discovery, drug discovery, food quality, etc. Indeed, the comprehensive non-biased identification and quantification of all the metabolites present in an organism, is a very ambitious goal, which, so far is not achieved. Due to the complexity and heterogeneity of physical and chemical properties of the metabolites, the idea of metabolomics is many times reduced to a more classical targeted approach. The dynamic characterizing the entire metabolome of a living organism implies that the data must cover a wide range of differences in space and in time. The current challenge of metabolomics is the measurement of the entire set of metabolome in a dynamic film, instead of in single snapshots.

Moreover, metabolomics should be complementary to transcriptomics and proteomics, which means that they must be integrated. Linking the different datasets of the entire systems biology platform will ensure a deeper insight into metabolic networks, and the understanding of cellular processes. But in one basic aspect metabolomics and other functional genomic information differ: the reproducibility of the data. Metabolomics data, mainly because of the lack of suitable analytical methods, cannot be stored in a permanent database or shared by different laboratories in different parts of the world. Furthermore, one standard metabolomic protocol has not been established yet. Much attention must be paid in all the steps of the analysis, from the preanalytical phase (sampling material, including harvesting, storing, extracting), through the analytical systems up to the data handling, in order not to include misleading data. Reproducibility, together with protocol standardization, must be considered as one of the primary tasks of current metabolomic techniques development. Only when protocols, libraries, and databases produced by different laboratories using different techniques will become publicly available, the metabolomics field will be able to match with the other "omics".

References

Afman L, Muller M (2006) Nutrigenomics: from molecular nutrition to prevention of disease. J Am Diet Assoc 106:569–576

Bednarek P, Schneider B, Svatos A et al. (2005) Structural complexity, differential response to infection and tissue specificity of indolic and phenylpropanoid secondary metabolism in *Arabidopsis* roots. Plant Physiol 138:1058–1070

Berrueta LA, Alonso-Salces RM, Heberger K (2007) Supervised pattern recognition in food analysis. J Chromatogr A 1158:196–214

Bino RJ, Hall RD, Fiehn O et al. (2004) Potential of metabolomics as a functional tool. Trends Plant Sci 9:418–425

Brescia MA, Caldarola V, de Giglio A et al. (2002) Characterization of the geographical origin of Italian red wine based on traditional and nuclear magnetic resonance spectrometric determinations. Anal Chim Acta 458:117–186

Brown SC, Kruppa G, Dasseux JL (2005) Metabolomics applications of FT-ICR mass spectrometry. Mass Spectrom Rev 24:223–231

Cardoso-Taketa AT, Pereda-Miranda R, Choi YH et al. (2008) Metabolomic profiling of the Mexican anxiolytic and sedative plant *Galphimia glauca* using nuclear magnetic resonance and multivariate data analysis. Planta Medica 74:1295–1301

Catchpole GS, Beckmann M, Enot DP et al. (2005) Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. Proc Natl Acad Sci U S A 102:14458–14462

Charlton A, Allnutt T, Holmes S et al. (2004) NMR profiling of transgenic peas. Plant Biotech J 2:27–35

Charlton AJ, Farrington WHH, Brereton P (2002) Application of ^{1}H-NMR and multivariate statistics for screening complex mixtures: quality control and authenticity of instant coffee. J Agric Food Chem 50:3098–3103

Choi YH, Tapias EC, Kim HK et al. (2004a) Metabolic discrimination of *Catharanthus roseus* leaves infected by phytoplasma using ^{1}H-NMR spectroscopy and multivariate data analysis. Plant Physiol 135:2398–2410

Choi YH, Kim HK, Hazekamp A et al. (2004b) Metabolomic differentiation of *Cannabis sativa* cultivars using ^{1}H NMR spectroscopy and principal component analysis. J Nat Prod 67:953–957

Choi HY, Kim HK, Linthorst HJM et al. (2006) NMR metabolomics to revisit the tobacco mosaic virus infection in *Nicotiana tabacum* leaves. J Nat Prod 69:742–748

Colquhoun IJ (2007) Use of NMR for metabolic profiling in plant systems. J Pestic Sci 32:200–212

Colquhoun IJ, Le Gall G, Elliott KA et al. (2006) Shall I compare thee to a GM potato? Trends Genetics 22:525–528

Cook D, Fowler S, Fiehn O, Thomashow MF (2004) A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of Arabidopsis. Proc Natl Acad Sci U S A 101:15243–15248

Dixon RA, Gang DR, Charlton AJ et al. (2006) Applications of metabolomics in agriculture. J Agric Food Chem 54:8984–8994

Dugo P, Kumm T, Crupi ML et al. (2006) Comprehensive two-dimensional liquid chromatography combined with mass spectrometric detection in the analyses of triacylglycerols in natural lipidic matrixes. J Chromatogr A 1112:269–275

Dunn WB, Ellis DI (2005) Metabolomics: current analytical platforms and methodologies. Trends Anal Chem 24:285–294

Eriksson L, Johansson E, Antti H et al. (2005) Multi- and megavariate data analysis: finding and using regularities in metabonomics data. In: Robertson DG, Lindon J, Nicholson JK, Holmes E (eds) Metabonomics in toxicity assessment. CRC, Boca Raton

Eriksson L, Johansson E, Kettaneh-Wold N et al. (2006) Multi- and megavariate data analysis. Umetrics AB, Umea, Sweden

Fiehn O (2001) Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. Comp Funct Genom 2:155–168

Fiehn O (2002) Metabolomics- the link between genotypes and phenotypes. Plant Mol Biol 48:155–171

Frederich M, Choi YH, Verpoorte R (2003) Quantitative analysis of strychnine and brucine in *Strychnos nux-vomica* using ^{1}H NMR. Planta Med 69:1169–1171

Fridman E, Pichersky E (2005) Metabolomics, genomics, proteomics, and the identification of enzymes and their substrates and products. Curr Opin Plant Biol 8:242–248

Fukusaki E, Kobayashi A (2005) Plant metabolomics: potential for practical operation. J Biosci Bioeng 4:347–354

Gibney MJ, Walsh M, Brennan L et al. (2005) Metabolomics in human nutrition: opportunities and challenges. Am J Clin Nutr 82:497–503

Goodacre R (2005) Metabolomics – the way forward. Metabolomics 1:1–2

Goodacre R (2007) Metabolomics of a superorganism. J Nutrit 137:259S–266S

Goodacre R, Vaidyanathan S, Dunn WB et al. (2004) Metabolomics by numbers: acquiring and understanding global metabolite data. Trends Biotechnol 22:245–252

Goodacre R, Broadhurst D, Smilde AK, et al. (2007) Proposed minimum reporting standards for data analysis in metabolomics. Metabolomics 3:231–241

Haas BJ, Wortman JR, Ronning CM et al. (2005) Complete reannotation of the *Arabidopsis* genome: Methods, tools, protocols and the final release. BMC Biol 3:7–27

Hall RD (2006) Plant metabolomics: from holistic hope, to hype, to hot topic. New Phytol 169:453–468

Harrigan GG, Martino-Catt S, Glenn KC (2007) Metabolomics, metabolic diversity and genetic variation in crops. Metabolomics 3:259–272

Hendrawati O, Yao Q, Kim HK et al. (2006) Metabolic differentiation of *Arabidopsis* treated with methyl jasmonate using nuclear magnetic resonance spectroscopy. Plant Sci 170:1118–1124

Hendriks MMWB, Cruz-Juarez L, De Bont D et al. (2005) Preprocessing and exploratory analysis of chromatographic profiles of plant extracts. Anal Chim Acta 545:53–64

Hirai MY, Yano M, Goodenowe DB et al. (2004) Integration of transcriptomics and metabolomics for understanding of global resonses to nutritional stresses in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A 101:10205–10201

Holmes E, Foxall D, Nicholson JK et al. (1994) Automatic data reduction and pattern recognition methods for analysis of ^{1}H nuclear magnetic resonance spectra of human urine from normal and pathological states. Anal Biochem 220:284–296

Hope JL, Prazen BJ, Nilsson EJ et al. (2005) Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry detection: analysis of amino acid and organic acid trimethylsilyl derivatives, with application to the analysis of metabolites in rye grass samples. Talanta 65:380–388

Jahangir M, Kim HK, Choi YH et al. (2007) Metabolomic response of *Brassica rapa* submitted to pre-harvest bacterial contamination. Food Chem 107:362–368

Keurentjes JJB, Fu J, de Vos CHR et al. (2006) The genetics of plant metabolism. Nature Genet 38:842–849

Kim HK, Choi YH, Erkelens C et al. (2005) Metabolic fingerprinting of *Ephedra* species using 1H NMR spectroscopy and principal component analysis. Chem Pharm Bull 53:105–109

Kim HK, Choi YH, Verpoorte R (2006) Metabolomic analysis of *Catharantus roseus* using NMR and principal component analysis in Plant Metabolomics, Springer, Heidelberg, Germany

Kruger NJ, Ratcliffe RG, Roscher A (2003) Quantitative approaches for analysing fluxes through plant metabolic networks using NMR and stable isotope labeling. Phytochem Rev 2:17–30

Kuiper HA, Kleter GA, Noteborn HPJM et al. (2002) Substantial equivalence-an appropriate paradigm for the safety assessment of genetically modified tools? Toxicol 181–182:427–431

Last RL, Jones AD, Shachar-Hill Y (2007) Towards the plant metabolome and beyond. Nat Rev Mol Cell Bio 8:167–174

Lay JO Jr, Borgmann S, Liyanage R et al. (2006) Problems with the "omics". Trends Anal Chem 25:1046–1056

Le Gall G, Puaud M, Colquihoun IJ (2001) Discrimination between orange juice and pulp wash by ^{1}H nuclear magnetic resonance spectroscopy: identification of marker compounds. J Agric Food Chem 49:580–588

Le Gall G, Colquhoun IJ, Davis AL et al. (2003) Metabolite profiling of tomato (*Lycopersicon esculentum*) using ^{1}H NMR spectroscopy as a tool to detect potential unintended effects following a genetic modification. J Agric Food Chem 51:2447–2456

Le Gall G, Metzdorff SB, Pedersen J et al. (2005) Metabolite profiling of *Arabidopsis thaliana* (L.) plants transformed with an antisense chalcone synthase gene. Metabolomics 1:181–198

Liang YS, Kim HK, Lefeber AWM et al. (2006) Identification of phenylpropanoids in methyl jasmonate treated *Brassica rapa* leaves using two-dimensional nuclear magnetic resonance spectroscopy. J Chromatogr A 1112:148–155

Lindon JC, Holmes E, Nicholson JK (2001) Pattern recognition methods and applications in biomedical magnetic resonance. Prog Nucl Mag Res Spectrosc 39:1–40

Lindon JC, Holmes E, Nicholson JK (2003) So what's the deal with metabonomics? Anal Chem 75:385A–391A

Manetti C, Bianchetti C, Bizzarri M et al. (2004) NMR-based metabonomic study of transgenic maize. Phytochemistry 65:3187–3198

Mariman ECM (2006) Nutrigenomics and nutrigenetics: the "omics" revolution in nutritional science. Biotechnol Appl Biochem 44:119–128

Markuszewski MJ, Britz-McKibbin P, Terabe S et al. (2003) Determination of pyridine and adenine nucleotide metabolites in *Bacillus subtilis* cell extract by sweeping borate complexation capillary electrophoresis. J Chromatogr A 989:293–301

Mondello L, Tranchida PQ, Stanek V et al. (2005) Silver-ion reversed phase comprehensive two-dimensional liquid chromatography combined with mass spectrometric detection in lipidic food analysis. J Chromatogr A 1086:91–98

Murch SJ, Rupasinghe HPV, Goodenowe DB et al. (2004) A metabolomic analysis of medicinal diversity in Huang- qin (*Scutellaria baicalensis*) genotypes: discovery of novel compounds. Plant Cell Rep 23:419–425

Mustafa NR (2007) A retrobiosynthetic study of salicylic acid production in *Catharantus roseus* cell suspension cultures. PhD thesis, Leiden University, Leiden

Nikiforova VJ, Gakiére B, Kempa S et al. (2004) Towards dissecting nutrient metabolism in plants: a system biology case study on sulphur metabolism. J Exp Bot 55:1861–1870

Noteborn HPJM, Lommena A, van der Jagta RC et al. (2000) Chemical fingerprinting for the evaluation of unintended secondary metabolic changes in transgenic food crops. J Biotechnol 77:103–114

Oliver SG, Winson MK, Kell DB et al. (1998) Systematic functional analysis of the yeast genome. Trends Biotechnol 16:373–378

Panagiotou G, Villas-Boas SG, Christakopoulos P et al. (2005) Intracellular metabolite profiling of *Fusarium oxysporum* converting glucose to ethanol. J Biotechnol 115:425–434

Pierce KM, Hope JC, Hoggard JC et al. (2006) A principal component analysis based method to discover chemical differences in comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (GC × GC-TOFMS) separations of metabolites in plant samples. Talanta 70:797–804

Ratcliffe RG, Schachar-Hill Y (2005) Revealing metabolic phenotypes in plants: inputs from NMR analysis. Biol Rev 80:27–43

Roos G, Roeseler C, Bueter KB et al. (2004) Classification and correlation of St. John's wort extracts by nuclear magnetic resonance spectroscopy, multivariate data analysis and pharmacological activity. Planta Med 70:771–777

Ryan D, Robards K (2006a) Metabolomics: the greatest omics of them all? Anal Chem 78:7954–7958

Ryan D, Robards K (2006b) Analytical chemistry considerations in plant metabolomics. Sep Purif Rev 35:319–356

Sato K, Soga T, Nishioka T et al. (2004) Simultaneous determination of the main metabolites in rice leaves using capillary electrophoresis mass spectrometry and capillary electrophoresis diode array detection. Plant J 40:151–163

Shellie RA, Marriott PJ, Huie CW (2003) Comprehensive two-dimensional gas chromatography (GC'GC) and GC'GC-quadrupole MS analysis of Asian and American ginseng. J Sep Sci 26:1185–1192

Shulaev V, Cortes D, Miller G et al. (2008) Metabolomics for plant stress response. Physiol Plantarum 132:199–208

Soga T, Ohashi Y, Ueno Y et al. (2003) Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. J Proteome Res 2:488–494

Sumner LW (2006) Current status and forward looking thoughts on LC/MS metabolomics. In: Saito K, Dixon R, Willmitzer L (eds) Biotechnology in agriculture and forestry: plant metabolomics. Springer, New York

Sumner LW, Mendes P, Dixon RA (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry 62:817–836

Tohge T, Nishiyama Y, Hirai MY et al. (2005) Functional genomics by integrated analysis of metabolome and trascriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. Plant J 42:218–234

van der Klift EJC, Vivó-Truyols G, Claassen FW et al. (2008) Comprehensive two-dimensional liquid chromatography with ultraviolet, evaporative light scattering and mass spectrometric detection of triacylglycerols in corn oil. J Chromatogr A 1178:43–55

Verpoorte R, van der Heijden R, ten Hoopen HJG et al. (1999) Metabolic engineering of plant secondary metabolite pathways for the production of fine chemicals. Biotechnol Lett 21:467–479
Verpoorte R, Choi YH, Kim HK (2005) Ethnopharmacology and system biology: a perfect holistic match. J Ethnopharmacol 100:53–56
Verpoorte R, Choi HC, Mustafa NR et al. (2008) Metabolomics: back to basics. Phytochem Rev 7:525–537
Vogels JTWE, Terwel L, Tas AC et al. (1996) Detection of adulteration in orange juices by a new screening method using proton NMR spectroscopy in combination with pattern recognition techniques. J Agric Food Chem 44:175–180
Wang Y, Tang H, Nicholson JK et al. (2004) Metabolomic strategy for the classification and quality control of phytomedicine: a case study of chamomile flower (*Matricaria recutita* L.) Planta Med 70:250–255
Watkins SM, German JB (2002) Toward the implementation of metabolomic assessments of human health and nutrition. Curr Opin Biotechnol 13:512–516
Weckwerth W (2003) Metabolomics in system biology. Annu Rev Plant Biol 54:669–689
Weckwerth W, Loureiro ME, Wenzel K et al. (2004) Differential metabolomic networks unravel the effects of silent plant phenotypes. Proc Natl Acad Sci U S A 101:7809–7814
Werner I, Bacher A, Eisenreich W (1997) Retrobiosynthetic NMR studies with ^{13}C-labeled glucose. J Biol Chem 272:25474–25482

Chapter 21
Transcriptomic Analysis of Multiple Enviornmental Stresses in Plants

Niranjani Jambunathan, Michael Puckette, and Ramamurthy Mahalingam

Abstract Global losses in agricultural production due to abiotic stresses have been estimated to be $120 billion. In the wake of shrinking arable land and rampant changes in climate, a second green revolution is important to meet the food needs of the rapidly growing population. This warrants radical changes in research strategies. Genomics approaches such as transcriptome profiling have led to the identification of gene networks important for many different stresses. However, under natural conditions plants are challenged by simultaneous occurrence of two or more stresses. Several studies have recently analyzed the transcriptional responses to two stresses simultaneously and are discussed here. For plant biotechnologies to deliver the promise of a second green revolution, a systems biology approach of examining multiple stresses at various developmental stages is necessary.

21.1 Introduction

Food security is a major issue in the global policy agenda. One of the targets of United Nations Millennium Development goal adopted in 2000 is to halve the proportion of people suffering from hunger by the year 2015 (The World Bank Group 2003). Meeting this food security goal is going to be a major challenge (Rosegrant and Cline 2003). In the next 40 years, demand for cereal production is predicted to increase by 60% as the population rises from the current 6.6 billion to 8.7 billion by the year 2050 (Bengtsson et al. 2006). In a world where population growth exceeds food supply (Malthus 1817), a second green revolution is necessary. Opportunities for plant biotechnologies to contribute to the second green revolution have been widely recognized (Fedoroff and Brown 2004).

N. Jambunathan, M. Puckette, and R. Mahalingam (✉)
Department of Biochemistry and Molecular Biology, Oklahoma State University, Stillwater, OK, 74078, USA
e-mail: ramamurthy.mahalingam@okstate.edu

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_21, © Springer Science+Business Media B.V. 2010

Three major yield-increasing strategies have been suggested – (1) Increasing harvest index [ratio of grain to total crop biomass], (2) increasing plant biomass, and (3) improving stress tolerance (Cassman 1999; Evans 1998). Nearly 82% of the potential yield of major crop losses occur every year due to abiotic stresses (Boyer 1982). Hence we consider improving stress tolerance in plants as the most important yield-increasing strategy. Further improving tolerance to stresses can also improve the harvest index, as well as biomass. Among the abiotic stresses, water deprivation or drought account for more than 50% of the estimated losses (Araus et al. 2002; Boyer 1982) and according to recent estimates accounts for more than $120 billion annually (Dhalmini et al. 2005). Plant water status is a central component of several abiotic stresses including drought, freezing and salinity (Verslues et al. 2006). In the wake of a global scarcity of water resources, drought stress is undoubtedly the major abiotic factor detrimental to agriculture.

The moisture status prevailing in soil environment in turn depends on the ambient environmental conditions such as temperature, light, humidity. Recent reports indicate that increase in global surface temperature is a major indicator of global warming (Van Vuuren et al. 2008). This raise in mean global temperature is attributed to increases in the greenhouse gases such as carbon dioxide and air pollutants such as ozone that has been brought about by anthropogenic activities. Thus plants not only have to deal with scanty water resources but may also have to deal with increasing temperatures, carbon dioxide and ozone. Even more worrisome is the fact some of these environmental variables can occur simultaneously, or successively, or in different times within a growing season. This has been elegantly illustrated in the form of a stress matrix depicting the combinations of agriculturally important biotic and abiotic stresses taking into consideration two different stresses at a time (Mittler 2006).

21.2 Physiological Studies of Combined Stresses

Practical agriculturists have long known that in field conditions plants face several distinct abiotic stresses either simultaneously or at different times during the growing season (Tester and Bacic 2005). Indeed breeding for tolerance to a combination of abiotic stresses is a common strategy for developing new crop varieties. Physiological studies of simultaneous exposure to different stresses in different plant species have been reported. Interactive effects of drought and heat shock stress represent the conditions encountered by plants growing in arid regions (Mittler et al. 2001; Moffat 2002). Analysis of these two stresses in Arabidopsis and tobacco plants showed that combination of heat and drought affected plants differently from drought or heat stress applied individually with reference to photosynthesis, respiration, stomatal conductance and leaf temperature (Rizhsky et al. 2002, 2004). In a study on sorghum, a combination of heat and drought stress during late stages of development [booting and flowering] caused nearly 87% yield reduction while the same stress treatment during vegetative stages had no impact

on yields (Craufurd et al. 1993). In another study, a combination of ozone and drought stress in two Norway spruce clones had varying effects on biomass depending on the vigour of the clone. In the faster growing clone, ozone reduced the biomass but also protected the plants from drought stress (Karlsson et al. 1997). A combination of high light and ozone in *Phaseolus vulgaris* was reported to exacerbate the detrimental effects of the pollutant on photosynthesis (Guidi et al. 2000). Modelling studies on the interactive effects of drought and ozone using yield forecasting predicted a general drought-induced reduction of crop sensitivity to ozone and it varied between crops, regions and years (King 1988). These physiological studies clearly illustrate that the abiotic stress combinations represent a new stress scenario and is not just a sum of the two independent stresses. Hence researchers working on improving stress tolerance in plants need to expand their areas of study to include stress combinations.

21.3 Molecular Genetics and Genomics

It is well known that adaptive responses to environmental stresses in plants are coordinated by molecular networks (Vinocur and Altman 2005). These networks help in re-establishing homeostasis and repair damaged proteins and membranes (Wang et al. 2003). In contrast to plant resistance to biotic stresses, which is mostly monogenic, abiotic stresses are multigenic traits and hence pose a formidable challenge for engineering resistance. The popular candidate gene-by-gene approach using genes from signalling and regulatory pathways (Seki et al. 2003; Shinozaki et al. 2003), proteins conferring stress tolerance (Wang and Huang 2004), enzymes leading to production of structural and functional metabolites (Park et al. 2004; Rontein et al. 2002) have met with some success in improving stress tolerance (Fig. 21.1). However many of these studies were short-term treatments considering a single stress and may not represent stress tolerance of crops in field conditions (Vinocur and Altman 2005).

Recently, the genomics approaches have radically changed the view about engineering stress tolerance in plants. The ability to investigate changes in thousands of genes simultaneously using microarray technology has led to a flurry of papers in both model systems as well as crop plants, in response to several different stresses (Kreps et al. 2002; Mahalingam et al. 2005, 2006; Puckette et al. 2008; Rossel et al. 2002; Seki et al. 2001). This is based on the premise that the set of genes that are expressed in a cell determines the state of the cell, how it is built, what it is capable of doing and perhaps what it is not capable of doing. There is a tight link between the function of a gene product and its expression pattern. Each gene is expressed in specific cells under specific conditions, and this regulation of when, where, and how much is fine-tuned by the process of natural selection. Promoters of genes operate as molecular switches by regulating transcription of specific genes in response to input information from external factors and internal milieu of a cell. These features of regulation of gene expression at the level of

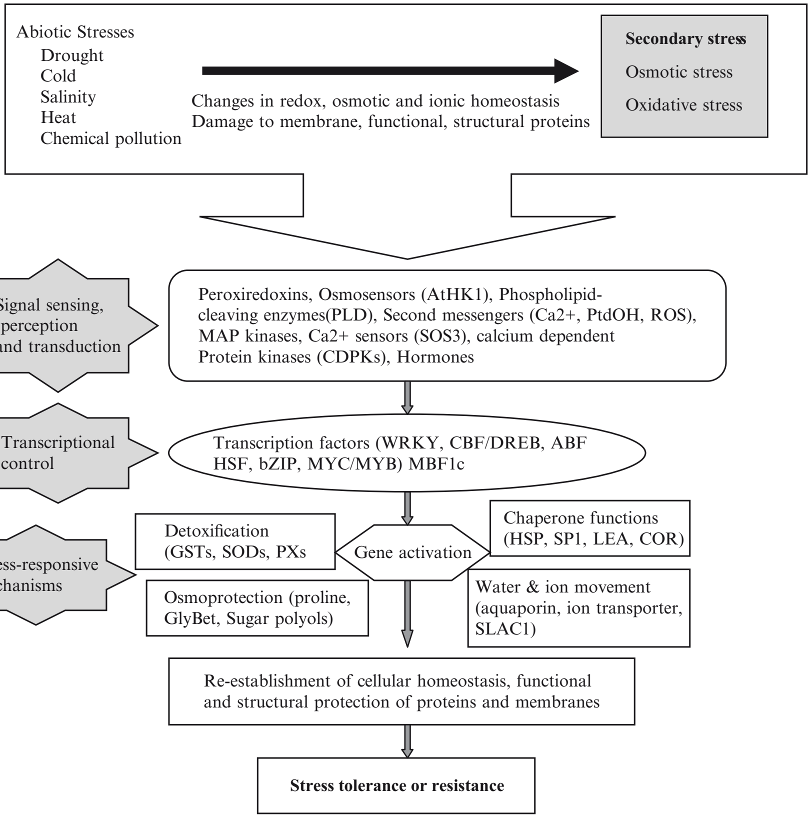


Fig 21.1 **A** general signal transduction pathway in response to abiotic stress in plants. Perturbations in the external environment leads to rapid changes in redox, osmotic and ionic balance in the plant cells. These changes are sensed by redox and osmosensors in conjunction with second messengers such as Calcium, ROS, RNS and phytohormones. These changes lead to activation of specific TFs which inturn leads to activating response genes that ultimately leads to stress tolerance or resistance. Inadequate responses at one or more steps in the signaling and gene activation process will result in irreversible changes in cellular homeostasis, lead to membrane and protein damage and ultimately cell death (Modified from Vinocur and Altman, 2005)

transcript abundance have been attributed to the wide popularity of microarray expression profiling (Brown and Botstein 1999).

Multiprotein Bridging Factor 1c (MBF1c) a common transcriptional coactivator recruits transcriptional machinery to transcription factor bound promoter (Naar et al. 2001). Based on the knowledge that transcription factors play a crucial role in regulating stress responses, MBF1c over-expressing plants were shown to give improved tolerance to environmental stresses (Suzuki et al. 2005).

The analysis of transcriptomes at various times following the stress treatment provides useful information about the temporal progression of thousands of genes.

In fact compendium of microarray data of entire Arabidopsis transcriptome in response to several stresses is easily accessible by programs like Genevestigator (Zimmermann et al. 2004). Using this compendium approach, a set of Multiple Stress Genes (MSTs) was identified in Arabidopsis (Kant et al. 2008). A subset of these genes belonging to the gene ontology category of transcription regulation was described as Multiple Stress Regulatory Genes (MSTR). Mutational analysis of several of the genes identified by this strategy confirmed the ability of these MSTR genes to confer resistance to multiple stresses (Kant et al. 2008). These studies again are examining a single stress at a time and suffer from the short-term nature of the treatments. As pointed out earlier, simultaneous occurrences of stresses is a common phenomenon in field conditions. Below we discuss several studies examining the transcriptomes following combined application of two stresses that are more likely to occur in realistic field conditions.

21.4 Analysis of Combined Stresses

21.4.1 Heat and High Light

Heat or high temperature stress adversely affect several physiological processes including photosystem II activity (Havaux et al. 1991), carbohydrate partitioning (Lafta and Lorenzen 1995), reactive oxygen metabolism (Larkindale et al. 2005), membrane fluidity (Kim and Portis 2005). Heat stress can also hamper growth processes (Beator and Kloppstech 1992; Cheikh and Jones 1994; Pressman et al. 2002; Wallwork et al. 1998). Excess light can be detrimental to photosynthesis, and lead to excess ROS (Niyogi 1999). Under natural conditions, a combination of high temperatures and high light are more likely to occur. Suppression subtraction hybridization technique was used to identify genes altered in response to combined heat and light stress in Arabidopsis (Nishizawa et al. 2006). Of the 76 genes identified in this study, was a heat shock transcription factor HsfA2. Constitutive overexpression of this gene led to enhanced resistance to combined heat, light and methyl viologen stress (Nishizawa et al. 2006).

In sunflower plants transcriptome analysis following high light, high temperature, and a combination stress showed that 105, 55 and 129 cDNAs were showing significant changes in gene expression, respectively (Hewezi et al. 2008). Significant differences were reported in the vegetative and storage tissues in combined stress treatment for the genes involved in photochemical reactions, carbon dioxide acquisition, pyruvate accumulation, lipid degradation, defense responses, protein folding and transcriptional gene regulation (Hewezi et al. 2008). This study also showed that the combined stress of heat and light induced the largest number of genes that were not altered when the stresses were applied singly (Fig. 21.2a). This study supports the notion that the combined heat and light stress represents a new type of stress.

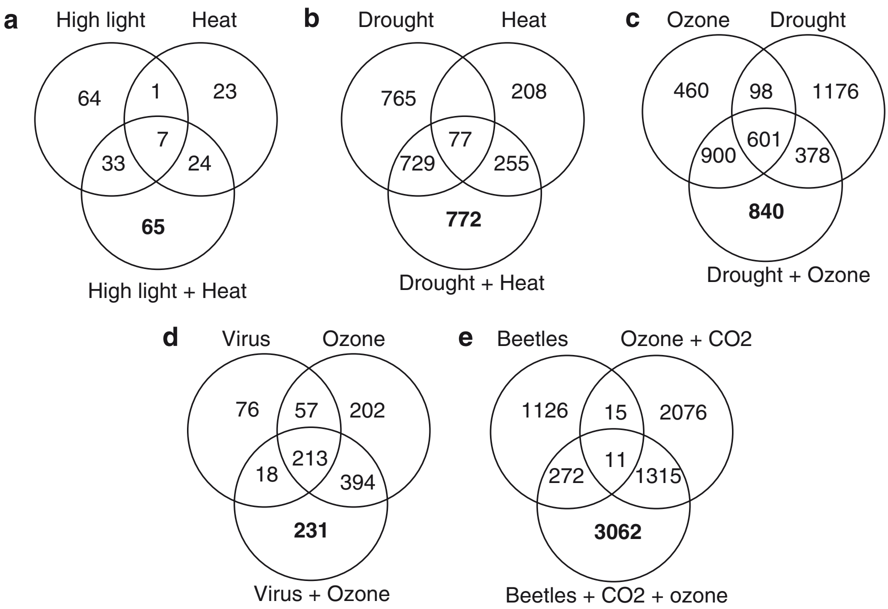


Fig. 21.2 Transcriptomic analysis of various combined stresses in plants. Figures shown here is summarized from data in a Hewezi et al., 2008; b Rizhsky et al., 2004; c Jambunathan et al., 2008; d Bilgin et al., 2008; e Casteel et al., 2008. Numbers in bold represent number of transcripts differentially expressed only in the combined stresses

21.4.2 Drought and Heat

Simultaneous drought and heat stress makes a case for a model combination stress. In a meta-analysis of all major US weather disasters between 1980 and 2004, it was reported that a combination of drought and heat led to loss of more than $120 billions in the United States alone (Mittler 2006). Drought and heat stress combination has been well studied in several crop plants and grasses (Craufurd et al. 1993; Heyne and Brunson 1940; Jagtap et al. 1998; Jiang and Huang 2001; Perdomo et al. 1996; Savage and Jacobson 1935; Savin and Nicolas 1996; Wang and Huang 2004). Physiological studies indicated that starch breakdown coupled with energy production in the mitochondria play a key role in plant metabolism during combined drought and heat stress (Rizhsky et al. 2002, 2004). Transcriptome analysis of drought, heat and combination of drought and heat indicated that the stress combination required more than 770 transcripts for acclimation and these were not altered by drought or heat stress (Rizhsky et al. 2004) (Fig. 21.2b).

21.4.3 Drought and Ozone

Ozone the most abundant air pollutant reduces plant biomass (Heagle 1989) by affecting allocation of assimilates and induces senescence process in plants (Miller et al. 1999; Pell et al. 1997). Highest ozone concentrations usually occur around mid-day and during summer season (Lorenzini et al. 1994). Field experiments on combined ozone and drought in Norway spruce revealed that ozone affected growth and biomass very differently in two different clones of this tree (Karlsson et al. 1997).

Since the flux of ozone into the plant is through the stomata, it has been argued that differences in sensitivity of plants to this pollutant are partly due to differences in stomatal conductance (Reich 1987). This dogma was challenged by elegant patch-clamping experiments to demonstrate that ozone directly affected guard cells by targeting inward potassium channels, which inhibited stomatal opening (Torsethaugen et al. 1999). Based on the ozone uptake models, drought induced stomatal closure will limit the ozone uptake into the leaves and hence protect the plants from ozone stress (Grunehage and Jager 2003; Panek and Goldstein 2001; Panek et al. 2002). However, this simplistic model has been confronted by the observation that ozone can cause stomatal "sluggishness" that lead to incomplete closure of the stomata and hence exacerbate the effects of drought (Grulke et al. 2003, 2005; Karnosky et al. 2005). The presence of nighttime ozone in rural locations (McCurdy 1994) can also greatly reduce the biomass of plants (Winner et al. 1989), especially if ozone can hinder the opening of the guard cells after a dark exposure (Torsethaugen et al. 1999). These studies show that impact of ozone and water deprivation together on plants is complex and merits attention given that concentrations of ozone in the troposphere will continue to increase in the future (Ashmore 2005).

In the model legume *Medicago truncatula* we have identified an accession JE154, which is resistant to a combination of chronic ozone and drought stress (Puckette et al. 2007). In the same study we also identified that the popular commercial cultivar Jemalong is sensitive to ozone and drought stress singly and in combination. Temporal transcriptome analysis in response to ozone in these two lines showed more than 2,000 genes were differentially expressed (Puckette et al. 2008). A striking finding was the rapid transcriptional reprogramming mounted by the resistant JE154 in stark contrast to the delayed response in sensitive Jemalong (Puckette et al. 2008). Recently we have examined the combination of chronic ozone and water deprivation stress in Jemalong using the Affymetrix gene chip technology. This analysis revealed that a combination of ozone and water stress simultaneously evoked a massive transcriptional reprogramming when compared with ozone or drought stress alone (Jambunathan et al. 2008) (Fig. 21.2c).

21.4.4 Carbon Dioxide and Ozone

Rising atmospheric [CO_2] and [O_3] have been recognized as important to crop production for only the last 30 or 40 years. The [CO_2] levels are reported to increase from the current levels of 380 umol mol^{-1} to nearly 550 umol. mol^{-1} by the year

2050 (IPCC 2007). Tropospheric O_3 levels in industrialized countries are currently 60 nmol. mol^{-1} and is predicted to increase by another 20% by 2050 (IPCC 2007). The most important concern related to improving plant tolerance to global climate change factors is that these are complex quantitative traits and hence are more difficult to control and engineer. To make the matters worse, ozone concentration is spatially and temporally heterogeneous owing to its short life and the fact its synthesis is dependent on the abundance of its pollutant precursors, sunlight and water vapour. The physiological consequences of elevated CO_2 and O_3 have been succinctly discussed in a recent review (Ainsworth et al. 2008).

Transcriptome analysis in aspen trees following long-term exposure to elevated CO_2 and tropospheric ozone, revealed that 238 genes were consistently differentially expressed in years 2001 and 2002 (Gupta et al. 2005). In the elevated CO_2 treatment, small number of genes was altered while a combination of both gases resulted in maximum number of changes in transcript levels. Transcripts related to photosynthesis were up regulated in the CO_2 treatment and was in support of the physiological studies showing elevated photosynthetic rates in these Aspen clones (Noormets et al. 2001). The elevated expression of xyloglucan endotransglycosylase responsible for cell wall loosening and wall expansion (Fry et al. 2002) may contribute to larger leaf size in response to increased CO_2 (Ferris et al. 2001). Downregulation of drought-induced aquaporin gene was taken as evidence for better water management in elevated CO_2 conditions. Elevated ozone alone leads to the down regulation of photosynthesis related genes and has been reported in several plant species (Mahalingam et al. 2005, 2006; Puckette et al. 2008). Under the combined treatment, a number of photosynthesis related genes were down regulated indicating that elevated CO_2 was not able to ameliorate the effects of elevated O_3. Again, the combined treatment of the two global climate change gases led to differential expression of a larger number of transcripts compared with either of them singly (Gupta et al. 2005).

21.4.5 *Ozone and Biotic Stress Interactions*

Ozone enters the plant through the stomata and are rapidly broken down into various reactive oxygen species at the apoplast (Kangasjarvi et al. 2005). By mimicking the oxidative burst that is observed in response to avirulent pathogens, plants sensitive to ozone develop hypersensitive response like localized lesions (Puckette et al. 2007; Rao and Davis 2001; Wohlgemuth et al. 2002). The ozone-induced lesions are regulated by the phytohormones such as salicylic acid, jasmonic acid, ethylene and abscisic acid (Ahlfors et al. 2004; Kangasjarvi et al. 2005; Overmyer et al. 2003; Rao et al. 2000, 2002). The changes in phytohormone levels in conjunction with the ROS levels leads to massive reprogramming of the transcriptome in response to elevated ozone (Li et al. 2006; Mahalingam et al. 2005, 2006; Puckette et al. 2008; Tosti et al. 2006). Some of the commonly induced transcripts in different plant species in response to elevated ozone include those belonging to the

category of stress/defense signaling. This begs the question of what happens to plant pathogen interactions in the wake of elevated ozone? In a study of soybean mosaic virus (SMV) in the presence of elevated levels of ozone it was reported that increased level of this pollutant reduced the systemic spread of the virus and reduced the overall disease development (Bilgin et al. 2008). This was attributed to the non-specific defence response caused by elevated levels of pathogenesis-related genes such PR-1, PR-5, PR-10 and EDS-1 and genes involved in flavonoid biosynthesis pathways. Overall the combined ozone and virus stress led to differential expression of larger number of transcripts compared to ozone or virus alone as the stresses progressed for a longer time (Fig. 21.2d).

21.5 Future Directions

From the above studies on examining two stresses jointly it is very clear that the combined stresses represents a new scenario to the plant and is simply not the sum of the two stresses. Secondly, large-scale transcriptomic studies have shown that in every case analysed to date, the combined stress treatments evoked the differential expression of larger number of transcripts compared to stresses applied singly. This again ratifies the observation that plants perceive combined stresses as something different. In the wake of these studies it is clear that genomics based approaches to improving stress tolerance in plants should move towards the next level of complexity. This entails analysis of multiple stresses simultaneously to simulate the conditions that usually operate in the field conditions. In fact, such a study has recently been reported wherein the impact of elevated CO_2 and elevated O_3 together was analysed for their impact on the resistance of soybean plants to the Japanese beetles (Casteel et al. 2008) (Fig. 21.2e). The combination of increased ozone and CO_2 along with the beetles, altered the transcript levels of more than 3,000 genes, elevated ozone and CO_2 changed the levels of 2,000 genes, while the beetle damage singly altered 1,100 genes (Casteel et al. 2008). Large scale transcriptomics data certainly provide a snap shot of the changes in the steady-state mRNA levels. Recent studies have indicated that post-transcriptional gene regulation plays a crucial role in plant stress signalling (Liu et al. 2008; Shukla et al. 2008; Sunkar et al. 2007). The dawn of the genomics era has led to emphasis on integrated analysis of stress response in whole plants tying together physiological and phenotypic observations with information on gene complement, transcript changes, protein complexes and metabolite levels (Bohnert et al. 2006). Combining transcriptomics of cell-and development stage specific profiling in roots in response to abiotic stress, it was elegantly demonstrated that transcriptional state of a cell is predominantly a reaction to environmental conditions regulated by a core set of genes that determines the cell identity (Dinneny et al. 2008). Thus it is apparent that to engineer tolerance to multiple traits that operate under realistic field conditions, plant molecular biologists and plant biotechnologists have to embrace the systems biology approaches to integrate data from multiple platforms.

Acknowledgements The authors thank Dr. Yuhong Tang, Samuel Roberts Noble Foundation for her assistance with the microarray analysis of combined ozone and drought stress in Medicago truncatula. Research in R.M. laboratory was supported by funding from USDA-NRI grant no. 2007-02635-0210696 and Oklahoma Agricultural Experiment Station.

References

Ahlfors R, Macioszek V, Rudd J, Brosche M, Schlichting R, Scheel D, Kangasjarvi J (2004) Stress hormone-independent activation and nuclear translocation of mitogen-activated protein kinases in *Arabidopsis thaliana* during ozone exposure. Plant J 40:512–522

Ainsworth EA, Rogers A, Leakey ADB (2008) Targets for crop biotechnology in a future high-CO2 and high-O-3 world. Plant Physiol 147:13–19

Araus JL, Slafer GA, Reynolds MP, Royo C (2002) Plant breeding and drought in C-3 cereals: What should we breed for? Ann Bot 89:925–940

Ashmore MR (2005) Assessing the future global impacts of ozone on vegetation. Plant Cell Environ 28:949–964

Beator J, Kloppstech K (1992) Heat-induced photomorphogenesis and circadian rhythmicity in messenger-RNA Levels and greening capacity in barley. Photosynth Res 34:182–182

Bengtsson M, Shen Y, Oki T (2006) A SRES-based gridded global population dataset for 1990–2100. Popul Environ 28:113–131

Bilgin DD, Aldea M, O'Neill BF, Benitez M, Li M, Clough SJ, DeLucia EH (2008) Elevated ozone alters soybean-virus interaction. Mol Plant Microbe Interact 21:1297–1308

Bohnert HJ, Gong QQ, Li PH, Ma SS (2006) Unraveling abiotic stress tolerance mechanisms – getting genomics going. Curr Opin Plant Biol 9:180–188

Boyer JS (1982) Plant Productivity and Environment. Science 218:443–448

Brown PO, Botstein D (1999) Exploring the new world of the genome with DNA microarrays. Nat Genet 21:33–37

Cassman KG (1999) Ecological intensification of cereal production systems: yield potential, soil quality, and precision agriculture. Proc Natl Acad Sci USA 96:5952–5959

Casteel CL, O'Neill BF, Zavala JA, Bilgin DD, Berenbaum MR, DeLucia EH (2008) Transcriptional profiling reveals elevated CO2 and elevated O-3 alter resistance of soybean (*Glycine max*) to Japanese beetles (*Popillia japonica*). Plant Cell Environ 31:419–434

Cheikh N, Jones RJ (1994) Disruption of maize kernel growth and development by heat-Stress – Role of cytokinin abscisic-acid balance. Plant Physiol 106:45–51

Craufurd PQ, Flower DJ, Peacock JM (1993) Effect of heat and drought stress on sorghum (*Sorghum bicolor*).1. Panicle development and leaf appearance. Aust J Exp Agric 29:61–76

Dhalmini Z, Spillane C, Moss JP, Ruane J, Urquia N, Sonnino A (2005) Status of research and application of crop biotechnologies in developing countries. Food and Agricultural Organization, Rome, pp 1–62

Dinneny JR, Long TA, Wang JY, Jung JW, Mace D, Pointer S, Barron C, Brady SM, Schiefelbein J, Benfey PN (2008) Cell identity mediates the response of *Arabidopsis* roots to abiotic stress. Science 320:942–945

Evans LT (1998) Feeding the ten billion: plant and population growth. Cambridge University Press, Cambridge

Fedoroff N, Brown NM (2004) Mendel in the Kitchen. Joseph Henry, Washington, DC

Ferris R, Sabatti M, Miglietta F, Mills RF, Taylor G (2001) Leaf area is stimulated in Populus by free air CO2 enrichment (POPFACE), through increased cell expansion and production. Plant Cell Environ 24:305–315

Fry SC, Miller JG, Dumville JC (2002) A proposed role for copper ions in cell wall loosening. Plant Soil 247:57–67

Grulke NE, Johnson R, Monschein S, Nikolova P, Tausz M (2003) Variation in morphological and biochemical O-3 injury attributes of mature Jeffrey pine within canopies and between microsites. Tree Physiol 23:923–929

Grulke NE, Dobrowolski W, Mingus P, Fenn ME (2005) California black oak response to nitrogen amendment at a high O-3, nitrogen-saturated site. Environ Pollut 137:536–545

Grunehage L, Jager H-J (2003) From critical levels to critical loads of ozone: a discussion of a new experimental and modeling approach for establishing flux-response relationships for agricultural crops and native plant species. Environ Pollut 125:99–110

Guidi L, Tonini M, Soldatini GF (2000) Effects of high light and ozone fumigation on photosynthesis in Phaseolus vulgaris. Plant Physiol Biochem 38:717–725

Gupta P, Duplessis S, White H, Karnosky DF, Martin F, Podila GK (2005) Gene expression patterns of trembling aspen trees following long-term exposure to interacting elevated CO2 and tropospheric O-3. New Phytol 167:129–142

Havaux M, Greppin H, Strasser RJ (1991) Functioning of Photosystem-I and Photosystem-II in Pea leaves exposed to heat stress in the presence or absence of light – Analysis using *in vivo* fluorescence, absorbency, oxygen and photoacoustic measurements. Planta 186:88–98

Heagle AS (1989) Ozone and crop yield. Annu Rev Phytopathol 27:397–412

Hewezi T, Leger M, Gentzbittel L (2008) A comprehensive analysis of the combined effects of high light and high temperature stresses on gene expression in sunflower. Ann Bot 102:127–140

Heyne EG, Brunson AM (1940) Genetic studies of heat and drought tolerance in maize. J Am Soc Agro 32:803–814

IPCC (2007) Climate change 2007: the physical science basis. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Contribution of Working Group I to the Fourth Annual Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, p 996

Jagtap V, Bhargava S, Streb P, Feierabend J (1998) Comparative effect of water, heat and light stresses on photosynthetic reactions in *Sorghum bicolor* (L.) Moench. J Exp Bot 49:1715–1721

Jambunathan N, Tang Y, Mahalingam R (2008) Analysis of combined chronic ozone and water deprivation stress in *Medicago truncatula*. Gordon Conference: Salt and water stress in plants. Big Sky, MT, USA

Jiang Y, Huang B (2001) Drought and heat stress injury to two cool season turfgrasses in relation to antioxidant metabolism and lipid peroxidation. Crop Sci 41:436–442

Kangasjarvi J, Jaspers P, Kollist H (2005) Signalling and cell death in ozone-exposed plants. Plant Cell Environ 28:1021–1036

Kant P, Gordon M, Kant S, Zolla G, Davydov O, Heimer YM, Chalifa-Caspi V, Shaked R, Barak S (2008) Functional genomics based identification of genes that regulate Arabidopsis responses to multiple abiotic stresses. Plant Cell Environ 31:697–714

Karlsson PE, Medin EL, Wallin G, Sellden G, Skarby L (1997) Effects of ozone and drought stress on the physiology and growth of two clones of Norway spruce (*Picea abies*). New Phytol 136:265–275

Karnosky DF, Pregitzer KS, Zak DR, Kubiske ME, Hendrey GR, Weinstein D, Nosal M, Percy KE (2005) Scaling ozone responses of forest trees to the ecosystem level in a changing climate. Plant Cell Environ 28:965–981

Kim K, Portis AR (2005) Temperature dependence of photosynthesis in Arabidopsis plants with modifications in Rubisco activase and membrane fluidity. Plant Cell Physiol 46:522–530

King DA (1988) Modeling the impact of Ozone X Drought interactions on regional crop yields. Environ Pollut 53:351–364

Kreps JA, Wu Y, Chang HS, Zhu T, Wang X, Harper JF (2002) Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress. Plant Physiol 130:2129–2141

Lafta AM, Lorenzen JH (1995) Effect of high temperature on plant growth and carbohydrate metabolism in Potato. Plant Physiol 109:637–643

Larkindale J, Hall JD, Knight MR, Vierling E (2005) Heat stress phenotypes of arabidopsis mutants implicate multiple signaling pathways in the acquisition of thermotolerance. Plant Physiol 138:882–897

Li PH, Mane SP, Sioson AA, Robinet CV, Heath LS, Bohnert HJ, Grene R (2006) Effects of chronic ozone exposure on gene expression in *Arabidopsis thaliana* ecotypes and in *Thellungielia halophila*. Plant Cell Environ 29:854–868

Liu CG, Calin GA, Volinia S, Croce CM (2008) MicroRNA expression profiling using microarrays. Nat Protoc 3:563–578

Lorenzini G, Nali C, Panicucci A (1994) Surface Ozone in Pisa (Italy)   a 6-year study. Atmos Environ 28:3155–3164

Mahalingam R, Shah N, Scrymgeour A, Fedoroff N (2005) Temporal evolution of the Arabidopsis oxidative stress response. Plant Mol Biol 57:709–730

Mahalingam R, Jambunathan N, Gunjan SK, Faustin E, Weng H, Ayoubi P (2006) Analysis of oxidative signalling induced by ozone in *Arabidopsis thaliana*. Plant Cell Environ 29:1357–1371

Malthus TR (1817) An essay on the principle of population. Murray, London

McCurdy TR (1994) Tropospheric ozone. In: McKee DJ (ed) Lewis Publishers, Boca Raton, pp 19–37

Miller JD, Arteca RN, Pell EJ (1999) Senescence-associated gene expression during ozone-induced leaf senescence in Arabidopsis. Plant Physiol 120:1015–1023

Mittler R (2006) Abiotic stress, the field environment and stress combination. Trends Plant Sci 11:15–19

Mittler R, Merquiol E, Hallak-Herr E, Rachmilevitch S, Kaplan A, Cohen M (2001) Living under a 'dormant' canopy: a molecular acclimation mechanism of the desert plant *Retama raetam*. Plant J 25:407–416

Moffat AS (2002) Plant genetics – Finding new ways to protect drought-stricken plants. Science 296:1226-–1229

Naar AM, Lemon BD, Tjian R (2001) Transcriptional coactivator complexes. Annu Rev Biochem 70:475–501

Nishizawa A, Yabuta Y, Yoshida E, Maruta T, Yoshimura K, Shigeoka S (2006) Arabidopsis heat shock transcription factor A2 as a key regulator in response to several types of environmental stress. Plant J 48:535–547

Niyogi KK (1999) Photoprotection revisited: Genetic and molecular approaches. Annu Rev Plant Physiol and Plant Mol Biol 50:333–359

Noormets A, McDonald EP, Dickson RE, Kruger EL, Sober A, Isebrands JG, Karnosky DF (2001) The effect of elevated carbon dioxide and ozone on leaf- and branch-level photosynthesis and potential plant-level carbon gain in aspen. Trees Struct Funct 15:262–270

Overmyer K, Brosche M, Kangasjarvi J (2003) Reactive oxygen species and hormonal control of cell death. Trends Plant Sci 8:335–342

Panek JA, Goldstein AH (2001) Response of stomatal conductance to drought in ponderosa pine: implications for carbon and ozone uptake. Tree Physiol 21:337–344

Panek JA, Kurpius MR, Goldstein AH (2002) An evaluation of ozone exposure metrics for a seasonally drought-stressed ponderosa pine ecosystem. Environ Pollut 117:93–100

Park EJ, Jeknic Z, Sakamoto A, DeNoma J, Yuwansiri R, Murata N, Chen THH (2004) Genetic engineering of glycinebetaine synthesis in tomato protects seeds, plants, and flowers from chilling damage. Plant J 40:474–487

Pell EJ, Schlagnhaufer CD, Arteca RN (1997) Ozone-induced oxidative stress: Mechanisms of action and reaction. Physiol Plantarum 100:264–273

Perdomo P, Murphy JA, Berkowitz GA (1996) Physiological changes associated with performance of Kentucky blue grass cultivars during summer stress. Hortscience 31:1182–1186

Pressman E, Peet MM, Pharr DM (2002) The effect of heat stress on tomato pollen characteristics is associated with changes in carbohydrate concentration in the developing anthers. Ann Bot 90:631–636

Puckette MC, Weng H, Mahalingam R (2007) Physiological and biochemical responses to acute ozone-induced oxidative stress in Medicago truncatula. Plant Physiol Biochem 45:70–79

Puckette MC, Tang YH, Mahalingam R (2008) Transcriptomic changes induced by acute ozone in resistant and sensitive Medicago truncatula accessions. BMC Plant Biol 8:46

Rao MV, Davis KR (2001) The physiology of ozone induced cell death. Planta 213:682–690
Rao MV, Lee H, Creelman RA, Mullet JE, Davis KR (2000) Jasmonic acid signaling modulates ozone-induced hypersensitive cell death. Plant Cell 12:1633–1646
Rao MV, Lee H, Davis KR (2002) Ozone-induced ethylene production is dependent on salicylic acid, and both salicylic acid and ethylene act in concert to regulate ozone-induced cell death. Plant J 32:447–456
Reich PB (1987) Quantifying plant response to ozone: a unifying theory. Tree Physiol 3:63–91
Rizhsky L, Liang H, Mittler R (2002) The combined effect of drought stress and heat shock on gene expression in tobacco. Plant Physiol 130:1143–1151
Rizhsky L, Liang H, Shuman J, Shulaev V, Davletova S, Mittler R (2004) When defense pathways collide. The response of Arabidopsis to a combination of drought and heat stress. Plant Physiol 134:1683–1696
Rontein D, Basset G, Hanson AD (2002) Metabolic engineering of osmoprotectant accumulation in plants. Metab Eng 4:49–56
Rosegrant MW, Cline SA (2003) Global food security: challenges and policies. Science 302:1917–1919
Rossel JB, Wilson IW, Pogson BJ (2002) Global changes in gene expression in response to high light in Arabidopsis. Plant Physiol 130:1109–1120
Savage DA, Jacobson LA (1935) The killing effect of heat and drought on buffalo grass and blue gamma grass at Hays, Kansas. J Am Soc Agro 27:566–582
Savin R, Nicolas ME (1996) Effects of short periods of drought and high temperature on grain growth and starch accumulation of two malting barley cultivars. Aust J Plant Physiol 23:201–210
Seki M, Narusaka M, Abe H, Kasuga M, Yamaguchi-Shinozaki K, Carninci P, Hayashizaki Y, Shinozaki K (2001) Monitoring the expression pattern of 1300 Arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray. Plant Cell 13:61–72
Seki M, Ishida J, Satou M, Sakurai T, Nakajima M, Enju A, Kamiya A, Akiyama K, Iida K, Narusaka M, Oono Y, Fujita M, Mizukado S, Nanjo T, Umezawa T, Kamei A, To T, Yamaguchi-Shinozaki K, Ecker JR, Davis RW, Theologis A, Carninci P, Kawai J, Hayashizaki Y, Sawasaki T, Kim JM, Hasegawa Y, Endo Y, Yokoyama S, Shinozaki K (2003) Arabidopsis encyclopedia using full-length cDNAs and its application for functional genomics. Plant Cell Physiol 44:S46–S46
Shinozaki K, Yamaguchi-Shinozaki K, Seki M (2003) Regulatory network of gene expression in the drought and cold stress responses. Curr Opin Plant Biol 6:410–417
Shukla LI, Chinnusamy V, Sunkar R (2008) The role of microRNAs and other endogenous small RNAs in plant stress responses. Biochim Biophys Acta
Sunkar R, Chinnusamy V, Zhu J, Zhu JK (2007) Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. Trends Plant Sci 12:301–309
Suzuki N, Rizhsky L, Liang HJ, Shuman J, Shulaev V, Mittler R (2005) Enhanced tolerance to environmental stress in transgenic plants expressing the transcriptional coactivator multiprotein bridging factor 1c. Plant Physiol 139:1313–1322
Tester M, Bacic A (2005) Abiotic stress tolerance in grasses. From model plants to crop plants. Plant Physiol 137:791–793
The World Bank Group (2003) Millennium development goals: about the goals. www.developmentgoals.org/About_the_goals.htm
Torsethaugen G, Pell EJ, Assmann SM (1999) Ozone inhibits guard cell K+ channels implicated in stomatal opening. Proc Natl Acad Sci USA 96:13577–13582
Tosti N, Pasqualini S, Borgogni A, Ederli L, Falistocco E, Crispi S, Paolocci F (2006) Gene expression profiles of O-3-treated Arabidopsis plants. Plant Cell Environ 29:1686–1702
Van Vuuren DP, Meinshausen M, Plattner GK, Joos F, Strassmann KM, Smith SJ, Wigley TM, Raper SC, Riahi K, de la Chesnaye F, den Elzen MG, Fujino J, Jiang K, Nakicenovic N, Paltsev S, Reilly JM (2008) Temperature increase of 21st century mitigation scenarios. Proc Natl Acad Sci USA 105:15258–15262
Verslues PE, Agarwal M, Katiyar-Agarwal S, Zhu JH, Zhu JK (2006) Methods and concepts in quantifying resistance to drought, salt and freezing, abiotic stresses that affect plant water status. Plant J 45:523–539

Vinocur B, Altman A (2005) Recent advances in engineering plant tolerance to abiotic stress: achievements and limitations. Curr Opin Biotechnol 16:123–132

Wallwork MAB, Jenner CF, Logue SJ, Sedgley M (1998) Effect of high temperature during grain-filling on the structure of developing and malted barley grains. Ann Bot 82:587–599

Wang WX, Vinocur B, Altman A (2003) Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. Planta 218:1–14

Wang ZL, Huang BR (2004) Physiological recovery of Kentucky bluegrass from simultaneous drought and heat stress. Crop Sci 44:1729–1736

Winner WE, Lefohn aS, Cotter IS, Greitner CS, Nellessen J, Mcevoy LR, Olson RL, Atkinson CJ, Moore LD (1989) Plant-responses to elevational gradients of O-3 exposures in Virginia. Proc Natl Acad Sci USA 86:8828–8832

Wohlgemuth H, Mittelstrass K, Kschieschan S, Bender J, Weigel HJ, Overmyer K, Kangasjarvi J, Sandermann H, Langebartels C (2002) Activation of an oxidative burst is a general feature of sensitive plants exposed to the air pollutant ozone. Plant Cell Environ 25:717–726

Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W (2004) GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. Plant Physiol 136:2621–263

Part IV
Transgenic Technologies

Chapter 22

Marker-Free Targeted Transformation

Hiroyasu Ebinuma and Kazuya Nanto

Abstract There are many transformation methods available for stable integration of a desirable gene into plant cells. In transformation, variable numbers of desired genes together with marker genes are randomly inserted into the plant genome. Therefore, cumbersome screening procedures are required to identify transgenic plants with a single copy of transgenes at appropriate expression levels. However, the lack of reproducibility of expression levels limits studies of both the gene expression and physiological effects of transgenes. And remaining of marker genes precludes retransformation with the same marker system and can raise safety and public concerns. The targeting approach is the best way to solve this problem. In this chapter, we focus on the application of site-specific recombination systems for introducing a desirable gene into a predefined site in a plant genome. Furthermore, we discuss an approach for removing a marker gene from targeted transgenic plants.

22.1 Introduction

With advances in gene recombination technology, the functional genome analysis of major crops has recently been vigorously promoted and several transgenic crops are now in large-scale commercial cultivation. For the commercialization of transgenic crops, added useful traits must be expressed stably and transmitted faithfully to the next generation. The problem must be overcome in order to select efficient reliable transgenic crops (Birch 1997).

Agrobacterium- and biolistic-mediated transformation methods are widely used to introduce a desirable gene into crops. However, even when the same gene is introduced into the same plant materials under the same culture conditions, transgenic crops still show large variations in transgene expression levels (Kohli et al. 2003). In transformation, a large number of genes are randomly inserted into the host's genome

H. Ebinuma (✉) and K. Nanto
R&D Division, Nippon Paper Industries Co. Ltd, Tokyo, Japan
e-mail: ebinumatokyo@yahoo.co.jp

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_22, © Springer Science+Business Media B.V. 2010

and the presence of excess copies of the transgene, tandem duplicated copies and partially deleted copies destabilizes transgene expression by RNA silencing (Matzke and Matzke 1991; Matzke and Mette 2000; Meyer 2000; Kumar and Fladung 2001). Since only a small number of transgenic plants have a single copy of stably expressed transgenes, a large number of transgenic plants must be produced and analyzed to select transgenic plants with a transgene at expected expression levels (Maqbool and Christou 1999; De Wilde et al. 2001; Vain et al. 2002). Moreover, it is impossible to repeatedly introduce a desirable gene into the same target site of a plant genome and to produce transgenic plants with a transgene at identical expression levels. This lack of reproducibility of transgene expression makes it difficult to compare close phenotypic differences among transgenic plants.

The targeting approach is the best way to solve this problem, by introducing a desirable gene into a predefined site in a plant genome (Ow 2002; Lyznik et al. 2003). With the aid of a site-specific recombination system, a site-specific integration method to insert a desirable gene into a predefined site (Albert et al. 1995; Vergunst and Hooykaas 1998; Vergunst et al. 1998; Day et al. 2000; Choi et al. 2000; Srivastava and Ow 2001; Srivastava et al. 2004; Chawla et al..2006) and the recombinase-mediated cassette exchange (RMCE) method to exchange a desirable gene with a target gene at a predefined site (Nanto et al. 2005; 2009; Nanto and Ebinuma 2008; Louwerse et al. 2007) are currently being developed. Systems for removing a marker gene from targeted transgenic plants with a transgene at a predefined site have recently been proposed and demonstrated (Srivastava and Ow 2004; Nanto and Ebinuma 2008). This report summarizes a targeting approach that includes a marker-free targeted transgenic plant.

22.2 Targeted Transformation Using a Site-Specific Recombination System

The introduction of a desirable gene into a predefined site in a plant genome by using a site-specific recombination system involves a two-stage process. First, with the use of conventional transformation methods, a target gene is randomly integrated into a plant genome and a transgenic plant with a single copy of stably expressed transgenes is selected as a target line. The target line has a target gene at a predefined target site in the plant genome. Next, a desirable gene is re-introduced into the target line and a transgenic plant with a desirable gene at a target site is selected.

To introduce a desirable gene into a target site, CRE/*lox* derived from bacteriophage P1 (Fukushige and Sauer 1992), FLP/*frt* from *Saccharomyces cerevisiae* (O'Gorman et al..1991) and R/*RS* from *Zygosaccharomyces rouxii* (Matsuzaki et al. 1990; Onouchi et al. 1991) are widely used. Site-specific recombination systems consist of two components: the recombinase and its recognition sequences. The recombinase efficiently mediates DNA recombination in two arranged recognition sequences. Two methods using site-specific integration and RMCE are designed to introduce a desirable gene into a target site in a plant genome. Figure 22.1a shows

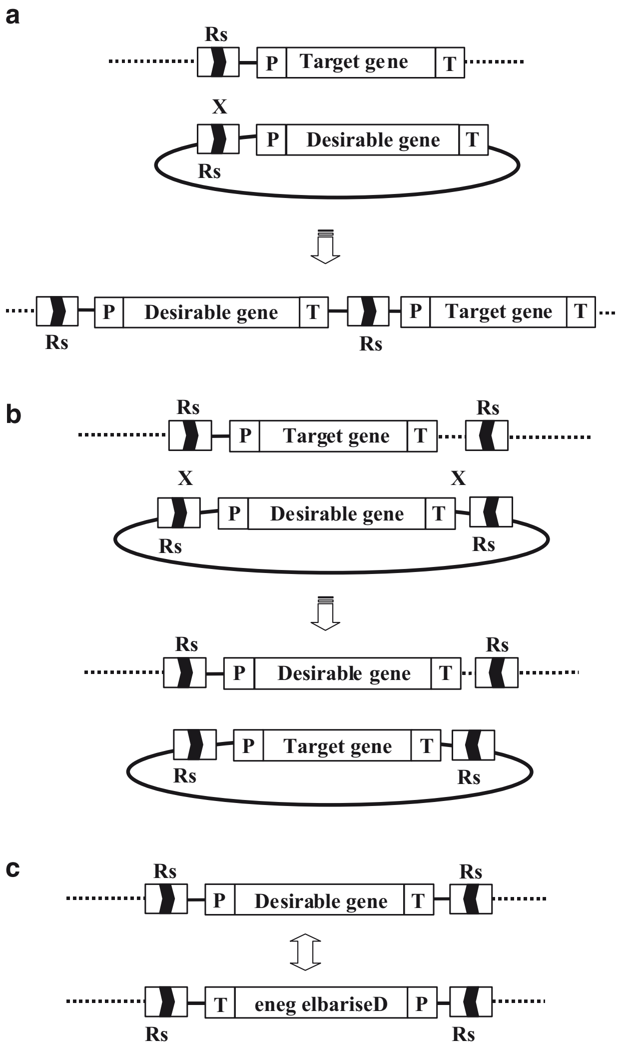


Fig. 22.1 Site-specific integration (**a**) and RMCE (**b**) methods. Genomic target sites containing one recognition sequence adjacent to a target gene (**a**) or two recognition sequences flanking a target gene in opposite orientation (**b**) are created by conventional transformation methods. Introduction of a construct carrying one recognition sequence adjacent to a desirable gene (**a**) or two recognition sequences flanking a desirable gene in opposite orientation (**b**) leads to targeted integration (**a**) or replacement via double reciprocal recombination (**b**). Since site-specific recombination reactions are fully reversible, recombinases catalyze excision of an integrated desirable gene flanked by two recognition sequences in direct orientation (**a**) but only inversion of an exchanged desirable gene flanked by two recognition sequences in opposite orientation (**c**). P: promoter, T: terminator, Rs: recognition sequence

a model for site-specific integration methods. The target line has a recognition sequence together with a target gene at a predefined target site in a plant genome. The plasmid DNA to re-introduce a desirable gene also has a recognition sequence. Since the recombinase mediates recombination in two recognition sequences, a desirable gene is inserted into a predefined target site in a plant genome. Figure 22.1b shows a model for RMCE methods. The target line has a target gene flanked by two recognition sequences in an opposite orientation at a predefined target site in a plant genome. The plasmid DNA has a desirable gene flanked by two recognition sequences in an opposite orientation. Since the recombinase mediates recombination in two recognition sequences, a desirable gene is exchanged with a target gene at a predefined target site in a plant genome.

22.3 Development of Targeted Transformation Methods

Targeted transformation methods using site-specific integration and RMCE have been reported in tobacco (Albert et al. 1995; Day et al. 2000; Choi et al. 2000; Nanto et al. 2005; 2009; Nanto and Ebinuma 2008), rice (Srivastava and Ow 2002; Srivastava et al. 2004; Chawla et al..2006) and *Arabidopsis* (Vergunst and Hooykaas 1998; Vergunst et al. 1998; Louwerse et al. 2007). Conventional transformation methods are used to re-introduce a desirable gene into target lines for site-specific integration and RMCE. Most of the desirable genes are randomly integrated into the plant genome of target lines, while the probability that a desirable gene will be introduced into a target site of a plant genome is extremely low. Therefore, it is very important to improve transformation processes for obtaining transgenic plants with a desirable gene at a target site. Several strategies have been employed with the aim of addressing the issues. Table 22.1 summarizes the properties of these targeted transformation methods.

22.3.1 Modification of Recognition Sequences

When a desirable gene is re-introduced by site-specific integration methods, the desirable gene integrated into a target site in a plant genome is flanked by two recognition sequences in a direct orientation, as shown in Fig. 22.1a. Since recombinase mediates the excision of a desirable gene from a target site due to its reversible reaction, it is difficult to stabilize the integration of a desirable gene into a target site. In tobacco (Albert et al. 1995; Day et al. 2000; Choi et al. 2000) and rice (Srivastava and Ow 2002; Srivastava et al. 2004; Chawla et al..2006), an approach to decrease the excision of a desirable gene from a target site by using mutant recognition sequences has been reported.

The recognition sequence of the CRE/*lox* recombination system consists of a spacer sequence of 8 bp flanked by two inverted repeat sequences of 13 bp as shown,

Table 22.1 Properties of targeted transformation methods

DNA delivery systems	***Agrobacterium*-mediated (b, c, h, j, k,l,m)**	**Biolistic-mediated (f, g, i)**	**Polyethylene glycol-mediated (a, d, e)**
Plant materials	**Tobacco (protoplasts: a, d, e) (leaf segments: h, j, k, l)**	***Arabidopsis* (root segments: b, c, m)**	**Rice (calluses: f, g, i)**
Site-specific recombination systems	**CRE/*lox* system (mutant recognition sites: a, b, c, d, e, f, g, i, m)**	**R/RS (h, j, k, l)**	
Targeted transformation methods (Figure 1, 2)	**Site-specific integration (mutant CRE/*lox*: a, b, c, d, e, f, g, i)**	**Recombinase-mediated cassette exchange (R/RS: h, j, k,l) (mutant CRE/*lox*: m)**	
Control of recombinase gene expression (Figure 3)	**Constitutive expression[1] (d, e, f, g, i)**	**Co-transformation[2] (a, b, c, m)**	**Auto-excision[3] (h, j, k, l)**
Selection of targeted gene (Figure 4)	**Silent marker[4] (a, b, c, d, e, f, g, i, m)**	**Auto-excision and negative marker[5] (h, k, l)**	**Negative marker[6] (j)**
Removal of randomly integrated copies (Figure 5)	**Segregation at next generation[7] (a, b, c, d, e, f, g, i, m)**	**Auto-excision and negative marker[8] (h, j, k, l)**	
Efficiency of targeted transformation methods[9]	**f: 1 plant / 2 plants / 36 callus plates, g: 12 plants / 3 plants / 79 callus plates, h: 7 plants / 6 plants / 128 leaf segments, j: 9 plants / 0 plants / 128 leaf segments, k: 18 plants / 0 plants / 576 leaf segments, l: 2 plants / 1 plant / 50 leaf segments, m: 1 plant / 3 plants / 13584 root segments.**		

[1] A recombinase gene is constitutively expressing at a target site of a plant genome. The expression is stopped by insertion of a desirable gene. [2]: One plasmid DNA contains a desirable gene and another one a recombinase gene. Two kinds of plasmid DNA are co-transformed. [3]: A plasmid DNA contains a recombinase gene flanked by two recognition sequences in direct orientation. Recombinase automatically excises it. [4]: A silent marker gene is activated by insertion into a target site. [5]: Positive and negative (*ipt*) marker genes are flanked by two recognition sequences in direct orientation. Recombinase automatically excises randomly integrated copies. [6]: A negative marker gene (*codA*) is placed at a target site of a plant genome. A desirable gene is exchanged with it by RMCE. [7]: A targeted desirable gene is segregated from randomly integrated copies at next generation. [8]: A desirable gene is flanked by two recognition sequences in direct orientation. Recombinase automatically excises randomly integrated copies. [9]: (the number of targeted transgenic plants with no additional copy/the number of targeted transgenic plants with additional copies/the number of explants) a: Albert et al. 1995, b: Vergunst and Hooykaas 1998, c: Vergunst et al. 1998, d: Day et al. 2000, e: Choi et al. 2000, f: Srivastava and Ow 2002, g: Srivastava et al. 2004, h: Nanto et al. 2005, i: Chawla et al. 2006, j: Ebinuma and Nanto 2007, k: Nanto et al. 2009, l: Nanto and Ebinuma 2008, m: Louwerse et al. 2007

in Fig. 22.2a. Each of two mutant recognition sequences has a mutation on one side of the inverted repeat sequences. Recombination between two mutant recognition sequences leads to wild-type recognition sequences and dual mutant recognition sequences that have a mutation on both sides of the inverted repeat sequences. The recombination frequency between wild-type and dual mutant recognition sequences is very low. In an improved site-specific integration method, plasmid DNA with one mutant recognition sequence together with a desirable gene is re-introduced into the target line that has another mutant recognition sequence together with a target gene, as shown in Fig. 22.2b. The desirable gene integrated into a target site is flanked by wild-type and dual mutant recognition sequences in a direct orientation. The low recombination between wild-type and dual mutant

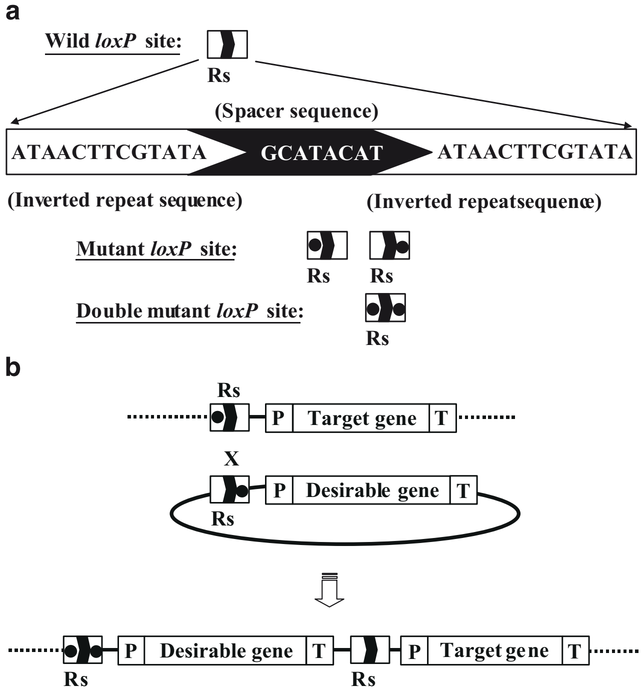


Fig. 22.2 Recognition sequences of the loxP site (**a**) and site-specific integration by using mutant loxP sites (**b**) (**a**) The *loxP* site consists of two 13 bp inverted repeats flanking an 8 bp spacer region. The mutant *loxP* site has a mutation on one side of the inverted repeat sequences, while the dual mutant *loxP* site has a mutation on both sides. (**b**) Introduction of a construct carrying one mutant *loxP* site into a target line containing another mutant *loxP* site leads to targeted integration of a desirable gene flanked by a wild *loxP* site and a dual mutant *loxP* site in direct orientation. Their low recombination frequency suppresses excision of an integrated desirable gene. P: promoter, T: terminator, Rs: recognition sequence

recognition sequences controls excision events and stabilizes the integration of a desirable gene into a target site in a plant genome.

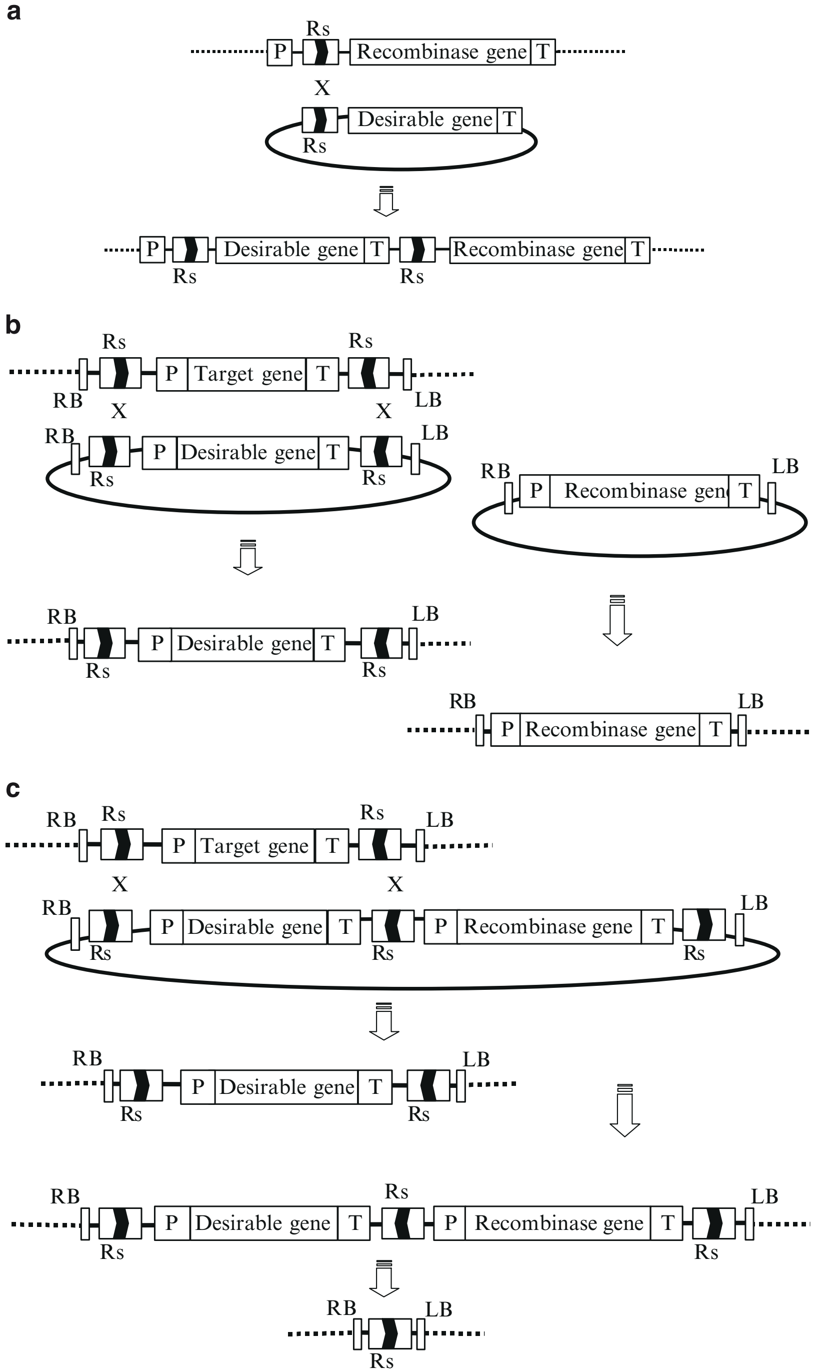
In the RMCE method, two recognition sites in an opposite orientation, as shown in Fig. 22.1b, flank the exchanged desirable gene with a target gene in a plant genome. Recombinase mediates inversion, rather than excision, of the exchanged desirable gene in a plant genome, as shown in Fig. 22.1c. Nanto et al. (2005) reportedly obtained targeted transgenic plants that have an exchanged desirable gene in direct and opposite orientations at a target site in a plant genome. Louwerse et al. (2007) reported the use of a wild-type recognition sequence and a mutant recognition sequence that has a mutation on a spacer sequence for RMCE. It is possible to obtain targeted transgenic plants with an exchanged desirable gene only in a direct orientation due to a lack of recombination between wild-type and mutant recognition sequences.

22.3.2 Control of Recombinase Gene Expression

In the introduction of a desirable gene into a target site by the use of a site-specific recombination system, recombinase mediates both integration and exchange events of a desirable gene. After the integration or exchange event of a desirable gene, recombinase falls into disuse and any remaining causes destabilization of a desirable gene at a target site in a plant genome. In site-specific integration methods using tobacco (Albert et al. 1995; Day et al. 2000; Choi et al. 2000) and rice (Srivastava and Ow 2002; Srivastava et al. 2004; Chawla et al..2006), a system in which the expression of a recombinase gene is stopped by integration of a desirable gene into a target site is adopted. The target line has a recombinase gene at a target site in a plant genome, as shown in Fig. 22.3a. The recombinase gene is expressed constitutively and has a recognition sequence between the promoter and coding sequences. A plasmid DNA that has a recognition sequence with a desirable gene is re-introduced into the target line. The plasmid DNA is inserted between the promoter and coding sequences of a recombinase gene and its expression is stopped.

Through the use of site-specific integration methods in tobacco (Albert et al. 1995) and in *Arabidopsis* (Vergunst and Hooykaas 1998; Vergunst et al. 1998), and RMCE methods in *Arabidopsis* (Louwerse et al. 2007) by the co-transformation of both the recombinase and desirable genes, a desirable gene is introduced into a target site. Two kinds of plasmid DNA are shown in Fig. 22.3b. One plasmid DNA has a recognition sequence together with a desirable gene and the other has a recombinase gene. The two kinds of plasmid DNA are introduced separately into *Agrobacterium*. By the co-transformation of two kinds of *Agrobacterium*, both kinds of plasmid DNA are introduced simultaneously into the same target line. Louwerse et al. (2007) reported obtaining targeted transgenic plants with a desirable gene at a target site in a plant genome. The targeted desirable gene at a target site and randomly integrated copies of recombinase genes could be segregated in the progeny of targeted transgenic plants. In some targeted transgenic plants, any recombinase gene remaining was not detected. These results show that a desirable gene can be introduced into a target site by the transient expression of a recombinase gene.

In RMCE methods with tobacco (Nanto et al. 2005; 2009; Nanto and Ebinuma 2008), the SDI system in which recombinase removes randomly integrated copies of recombinase genes from a plant genome is adopted. A plasmid DNA that has a recombinase gene flanked by two recognition sequences in direct orientation together with a desirable gene is re-introduced into a target line, as shown in Fig. 22.3c. A desirable gene is exchanged with a target gene at a target site in a plant genome through the mediation of a recombinase gene on the same plasmid DNA. The randomly integrated copies of a recombinase gene in a plant genome are removed by recombinase. Nanto et al. (2005) reported that none of the targeted transgenic plants had a recombinase gene and no footprint that would indicate the removal of a recombinase gene from a plant genome was detected in targeted transgenic plants. These results show that the SDI system can efficiently produce targeted transgenic plants with no recombinase gene and the transient expression of a recombinase gene can mediate the exchange of a desirable gene with a target gene.


a
Rs
P
Recombinase gene
T
X
Desirable gene
T
Rs
P
Rs
Desirable gene
T
Rs
Recombinase gene
T
b
Rs
Rs
P
Target gene
T
RB
LB
X
X
RB
LB
P
Desirable gene
T
Rs
Rs
RB
LB
P
Recombinase gen
T
RB
LB
P
Desirable gene
T
Rs
Rs
RB
LB
P
Recombinase gene
T
c
RB
Rs
Rs
LB
P
Target gene
T
X
X
RB
LB
P
Desirable gene
T
P
Recombinase gene
T
Rs
Rs
Rs
RB
LB
P
Desirable gene
T
Rs
Rs
RB
Rs
LB
P
Desirable gene
T
P
Recombinase gene
T
Rs
Rs
RB
LB
Rs

22.3.3 DNA Delivery Methods

Various DNA delivery methods can be used to re-introduce a plasmid DNA with a desirable gene into a target line. Double-stranded DNA is a good substrate for recombinase of a site-specific recombination system. With the use of PEG-mediated protoplast transformation (Albert et al. 1995; Day et al. 2000; Choi et al. 2000) and biolistic-mediated transformation (Srivastava and Ow 2002; Srivastava et al. 2004; Chawla et al..2006), where a large amount of double-stranded DNA is introduced into a plant cell, the successful production of targeted transgenic plants has been reported.

Meanwhile, with the use of *Agrobacterium*-mediated transformation (Vergunst and Hooykaas 1998; Vergunst et al. 1998), where a small amount of linear single-stranded T-DNA is introduced into a plant cell, very few targeted transgenic plants have been reported. Thus, T-DNA was believed to be a poor substrate for recombinase in a site-specific recombination system.

Recently, the mechanism by which T-DNA is integrated by *Agrobacterium* was revealed (Chilton and Que 2003; Tzfira et al. 2003). When *Agrobacterium* introduces T-DNA into a plant cell, it is converted to a linear double-stranded T-DNA intermediate, which is integrated into the plant genome. With the use of *Agrobacterium*-mediated transformation in tobacco (Nanto et al. 2005; 2009; Nanto and Ebinuma 2008) and *Arabidopsis* (Louwerse et al. 2007), a desirable gene of plasmid T-DNA has been reported to exchange with a target gene of a plant genome at a fairly high rate. In *Arabidopsis* (Louwerse et al. 2007), no additional copy of a desirable gene that was randomly integrated into the plant genome was detected in targeted transgenic plants, and in tobacco (Nanto et al. 2005; in press; Nanto and Ebinuma 2008), no footprint that would indicate the removal of a desirable gene that was randomly integrated into a plant genome was detected in targeted transgenic plants. These results show that linear single-stranded T-DNA is not a

Fig. 22.3 Control strategies of recombinase gene expression. (**a**) Genomic target sites contain a recognition sequence between the promoter and coding sequences of the recombinase gene which is expressed constitutively. Introduction of a construct carrying one recognition sequence leads to insertion of a desirable gene between the promoter and coding sequences. The recombinase gene expression is stopped. (**b**) Genomic target sites contain two recognition sequences flanking a target gene in opposite orientation. One construct carries two recognition sequences flanking a desirable gene in opposite orientation. Another construct carries only a recombinase gene. Co-introduction of both these constructs leads to replacement of a target gene with a desirable gene by transient expression of a recombinase gene. Randomly integrated copies of recombinase genes are segregated at the next generation. (**c**) Genomic target sites contain two recognition sequences flanking a target gene in opposite orientation. A construct carries two recognition sequences flanking a recombinase gene in direct orientation together with a desirable gene that is flanked by two recognition sequences in opposite orientation. Introduction of the construct leads to replacement of a target gene with a desirable gene by transient expression of a recombinase gene. Recombinases catalyze excision of randomly integrated copies of both the desirable and recombinase genes. P: promoter, T: terminator, Rs: recognition sequence, LB: left border, RB: right border

limiting substrate for recombinase of a site-specific recombination system and a desirable gene of a linear double-stranded T-DNA intermediate can exchange directly with a target gene in a plant genome.

22.3.4 Selection of Targeted Transgenic Plants

With the current transformation methods, a desired gene is randomly integrated into a plant genome. Since the proportion of targeted transgenic plants to randomly integrated ones is extremely low, the development of highly efficient selection methods has become a major issue. To overcome this problem, a strategy in which a selection marker gene can be activated by introducing a desirable gene into a target site is widely used (Albert et al. 1995; Vergunst and Hooykaas 1998; Vergunst et al. 1998; Day et al. 2000; Choi et al. 2000; Srivastava and Ow 2002; Srivastava et al. 2004; Chawla et al..2006; Louwerse et al. 2007). As shown in Fig. 22.4a, in site-specific integration methods, the target line has a transgene where a recognition sequence is placed between the promoter and coding sequences. A plasmid DNA with a recognition sequence and a silent marker gene without a promoter sequence is re-introduced into the target line. When it is integrated into a target site, the coding sequence of a marker gene is joined to the downstream region of a promoter sequence of a transgene at a target site and the expression of a marker gene is activated.

Nanto et al. (2005; 2009) reported the design of a system that can efficiently enrich gene exchange events through the removal of randomly integrated genes. As shown in Fig. 22.4b, a plasmid DNA has a desirable gene, a marker gene and a *ipt* gene that are flanked by two recognition sequences in a direct orientation. With the use of *Agrobacterium* methods, a plasmid DNA is re-introduced into a target line. For some plasmid DNA, both the desirable and marker genes are exchanged with a target gene at a target site. Recombinase mediates the excision of randomly integrated plasmids DNA from a plant genome. The *ipt* gene is used as a negative selection marker to identify transgenic plants that still contain randomly integrated plasmid DNA. Since transgenic plants with an *ipt* gene show a shooty phenotype without rooting ability due to the overproduction of cytokinin, they can be easily identified (Ebinuma et al. 1997; Sugita et al. 1999). In addition, this system can be expected to show an increased frequency of gene exchange events during transformation since cytokinin induces callus formation and proliferates selectively transgenic cells in which a plasmid DNA with desirable and recombinase genes are introduced.

22.4 Marker-Free Targeted Transgenic Plants

Conventional transformation methods widely use a marker gene such as an antibiotic-resistance gene to select a transgenic plant. A selection marker gene remaining in a transgenic plant interrupts the accumulation of transgenes by re-transformation

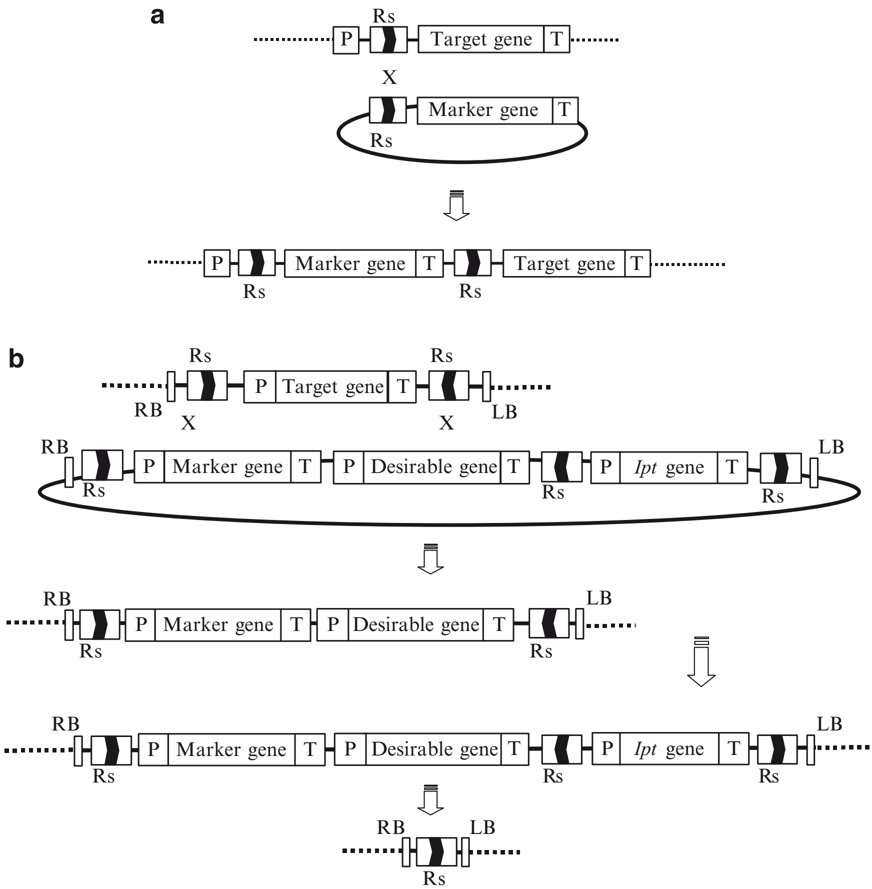


Fig. 22.4 Selection strategies of targeted transgenic plants. (**a**) Genomic target sites contain a recognition sequence between the promoter and coding sequences of the target gene that is expressed constitutively. A construct carries one recognition sequence adjacent to a silent marker gene without a promoter sequence. Introduction of the construct leads to insertion of a silent marker gene between the promoter and coding sequences of the target gene. The marker gene expression is activated. (**b**) Genomic target sites contain two recognition sequences flanking a target gene in opposite orientation. A construct carries two recognition sequences flanking an *ipt* gene in direct orientation together with marker and desirable genes that are flanked by two recognition sequences in opposite orientation. Introduction of the construct leads to replacement of a target gene with marker and desirable genes. Recombinases catalyze excision of randomly integrated copies of the marker, desirable and *ipt* genes. The *ipt* gene is used as a negative selection marker to eliminate transgenic plants where randomly integrated copies are remaining. P: promoter, T: terminator, Rs: recognition sequence, LB: left border, RB: right border

and reduces the acceptability of transgenic crops by consumers. Currently, the segregation of a marker gene by co-transformation systems and the removal of a marker gene by site-specific recombination systems have been reported to produce marker-free transgenic plants (Ebinuma et al. 2001, 2004, 2005). In a targeted transformation method, the introduction of both the desirable and marker genes into the same target site is needed to select a targeted transgenic plant. When marker-free transgenic plants are produced by co-transformation methods, the desirable and marker genes must both be separately introduced into sites on different chromosomes to segregate the desirable gene from the marker gene in the next generation. Therefore, co-transformation methods cannot be used to remove a marker gene from a targeted transgenic plant.

Recently, the removal of a marker gene from a targeted transgenic plant by the use of a site-specific recombination system has been reported (Nanto and Ebinuma 2007). The RMCE method is used to produce a targeted transgenic plant with desirable and marker genes at a target site in a plant genome. As shown in Fig. 22.5a, the marker gene of a targeted transgenic plant is flanked by two recognition sequences in a direct orientation. The plasmid DNA that has *ipt* and recombinase genes flanked by two recognition sequences in a direct orientation is re-introduced into the targeted transgenic plant. Recombinase mediates the removal of a marker gene at a target site and randomly integrated plasmid DNA from a plant genome. The *ipt* gene is used as a negative selection marker to exclude transgenic plants with randomly integrated plasmid DNA (Ebinuma et al. 1997; Sugita et al. 1999). The overproduction of cytokinin by the *ipt* gene regenerates transgenic plants in hormone-free medium and, based on a DNA analysis, half of the population with a normal phenotype becomes marker-free targeted transgenic plants. Since some of the marker-free targeted transgenic plants have no footprint that would indicate the removal of randomly integrated plasmid DNA from the plant genome, this result shows that the transient expression of a recombinase gene can efficiently remove a marker gene from a plant genome.

Meanwhile, Ebinuma and Nanto (2007) reported the selection of marker-free targeted transgenic plants by using a negative marker gene (*cod*A) with RMCE methods. The *cod*A gene codes for a cytosine deaminase, which converts 5-fluorocytosine (5-FC) to 5-fluorouracil (5-FU). Since 5-FU is toxic to plant cells, a transgenic plant with a codA gene is blighted on medium containing 5-FC (Schlaman and Hooykaas 1997). As shown in Fig. 22.5b, a target transgenic line has a *cod*A gene at a target site in either a heterozygous or homozygous condition. With the use of *Agrobacterium* methods, plasmid DNA with only a desirable gene is re-introduced into the target lines. *Agrobacterium* is infected to 64 pieces of tobacco leaf and transgenic plants that survive on medium containing 5-FC are selected. Among the transgenic plants that were resistant to 5-FC, 20% to 30% were shown to be marker-free targeted transgenic plants. Interestingly, the results show that marker-free transgenic plants with a desirable gene at a target site in a homozygous condition can be selected if target lines with a *cod*A gene at a target site in a homozygous condition are used for transformation.

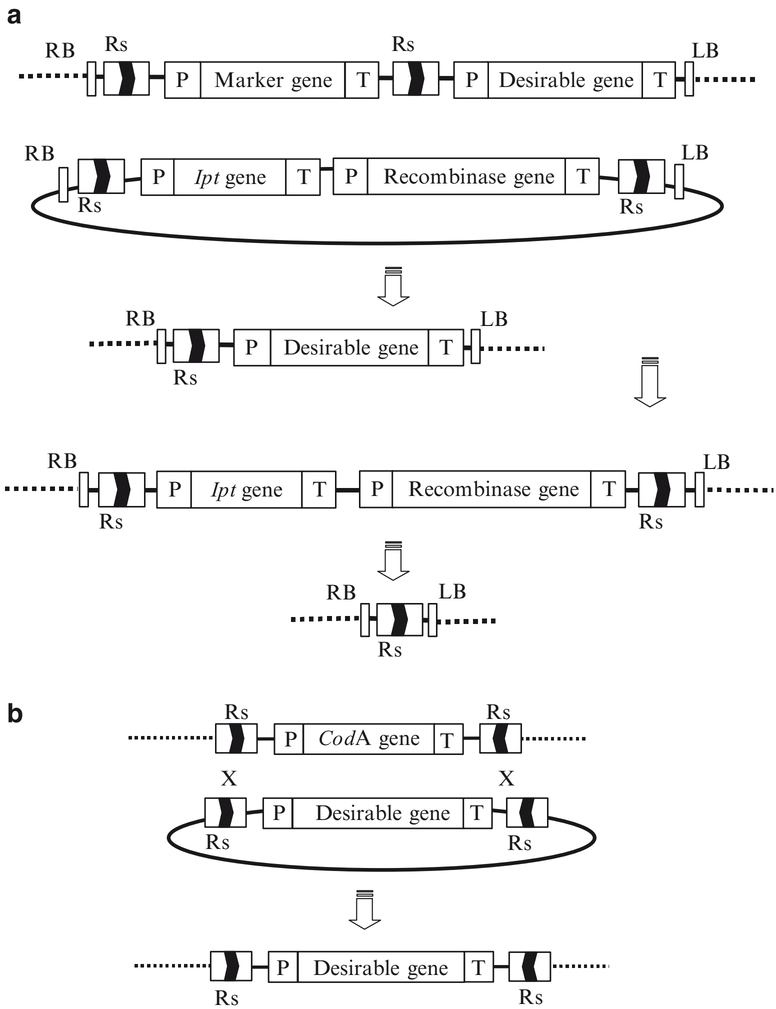


Fig. 22.5 Selection strategies of marker-free targeted transgenic plants. (**a**) Targeted transgenic plants have two recognition sequences flanking a marker gene in direct orientation adjacent to a desirable gene. Introduction of a construct carrying two recognition sequences flanking *ipt* and recombinase genes in direct orientation leads to excision of a marker gene at genomic target sites. Recombinases catalyze excision of randomly integrated copies of the *ipt* and recombinase genes. The *ipt* gene is used as a negative selection marker to eliminate transgenic plants where randomly integrated copies are remaining. (**b**) Genomic target sites contain two recognition sequences flanking a *codA* gene in opposite orientation. Introduction of a construct carrying two recognition sequences flanking a desirable gene in opposite orientation leads to replacement of a *codA* gene with a desirable gene. The *codA* gene is used as a negative selection marker to eliminate transgenic plants where *codA* genes are remaining. P: promoter, T: terminator, Rs: recognition sequence, LB: left border, RB: right border

22.5 Reproducibility of Transgene Expression

The so-called "position effect" refers to a difference in expression levels among transgenes at different positions in a plant genome. However, the technical difficulty of introducing a desirable gene into a predefined site in a plant genome has been a major issue in clarifying the position effect (Meyer 2000; Kumar and Fladung 2001; Koli et al. 2003; De Buck et al. 2004; Kim et al. 2007). Targeted transformation methods are useful tools to control variable transgene expression caused by transgene integration patterns and genomic location.

In cultured animal cells, the development of RMCE methods using a site-specific recombination system is underway and the effects of gene structures and chromosomal positions on the expression of transgenes have been examined in detail (Feng et al. 1999, 2001; Eszterhas et al. 2002). When a desirable gene is introduced into a predefined site with high transcriptional activity by the RMCE method, the gene expression levels are highly predictable and repeatable, with very little variation. The gene expression level clearly depends on the site at which a desirable gene is introduced and also the orientation of the introduced gene constructs. When two genes joined in a tandem, divergent or convergent orientation are introduced into a predefined site, the strongest interference with gene expression is detected with a convergent orientation and gene expression is most strongly affected by chromosomal positions in a tandem orientation.

In rice, Chawla et al. (2006) reported to obtain targeted transgenic plants that contain a single copy of desirable genes at a target site but not additional copies at random genomic sites by using the site-specific integration method. The desirable genes at target sites were expressed at consistent levels over three to four generations and the gene expression levels nearly doubled in the homozygous progeny. In tobacco, Nanto et al. (2009) examined the effect of chromosomal position on gene expression by using the RMCE method. A desirable gene is re-introduced into three different target lines of tobacco by the RMCE method and into a control line by the conventional method. The variation of gene expression levels within desirable genes at the same target site is very slight compared with that among randomly integrated genes. There is a large difference in gene expression levels for a desirable gene at three target sites, but there is no correlation of gene expression levels between a target gene at three target sites and a desirable gene exchanged with the target gene. Interestingly, when one, two and three desirable genes joined in a tandem orientation are re-introduced into the same target line (Nanto and Ebinuma 2007), the gene expression levels increase more than two- and threefold. The orientation of desirable genes at a target site has been shown to influence their expression.

22.6 Conclusion

In the context of global climate changes and rapidly growing food demands, development of reliable transformation methods is a major target for trait improvement of commercial crops. Since plant transformation is a fundamental technique in the

functional analysis of genes as well as in the production of transgenic crops. In transformation, variable numbers of desired genes together with marker genes are randomly inserted into the plant genome. Therefore, it requires a large amount of DNA and RNA analysis to identify a transgenic plant with a single copy of a desired gene with a high expression level. However, the lack of reproducibility of expression levels limits studies of both the gene expression and physiological effects of transgenes. And remaining of marker genes precludes retransformation with the same marker system and can raise safety and public concerns.

Compared with conventional methods, the RMCE method enables us to easily identify such reliable transgenic plants simply by using a PCR analysis, and to consistently reproduce the same gene expression level. Furthermore, marker-free targeted transgenic plants can be selected by using a negative marker gene (*codA*). If we use the same target lines as common standard lines, it is possible to compare gene expression levels at the same target site in a plant genome. Such a common evaluation criterion is very useful for precisely investigating gene characteristics and genome functions, and for optimizing a gene structure for research and commercial use based on the high predictability and repeatability of gene expression.

In a negative selection system, the target line has a single copy of a *codA* gene with consistent expression levels. By using a target line with high sensitivity to 5-FC, high selection efficiency could be achieved. Instead of a *codA* gene, visual reporter genes and regeneration promotion genes could be used to promote both the regeneration and identification of transgenic cells during the transformation process. If a target line for the RMCE method has such suitable characteristics for transformation, transformation efficiency can increase dramatically in recalcitrant plant species.

References

Albert H, Dale E-C, Lee E et al. (1995) Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. Plant J 7:649–659

Birch R-G (1997) Plant transformation: problems and strategies for practical application. Annu Rev Plant Physiol Plant Mol Biol 48:297–326

Chawla C, Ariza-Nieto M, Wilson A-J et al. (2006) Transgene expression produced by biolistic-mediated site-specific gene integration is consistently inherited by the subsequent generations. Plant Biotechnol J 4:209–218

Chilton M-D, Que Q (2003) Targeted integration of T-DNA into the tobacco genome at double-strand breaks: new insights on the mechanism of T-DNA integration. Plant Physiol 133:956–965

Choi S, Begum D, Koshinsky H et al. (2000) A new approach for the identification and cloning of genes: the pBACwich system using Cre/lox site-specific recombination. Nucleic Acids Res 28:E19–E26

Day C-D, Lee E, Kobayashi J et al. (2000) Transgene integration into the same chromosome location can produce alleles that express at a predictable level, or alleles that are differentially silenced. Genes Dev 14:2869–2880

De Buck S, Windels P, De Loose M et al. (2004) Single-copy T-DNAs integrated at different positions in the *Arabidopsis* genome display uniform and comparable B-glucuronidase accumulation levels. CMLS Cell Mol Life Sci 61:2632–2645

De Wilde C, Podevin N, Windels P et al. (2001) Silencing of antibody genes in plants with single-copy transgene inserts as a result of gene dosage effects. Mol Genet Genom 265:647–653

Ebinuma H, Nanto K (2007) Marker-free transgenic plants with a site-specific integrated copy produced by recombinase-mediated cassette exchange (RMCE). International Conference – Plant Tansformation Technologies (Abstract: pp 41). Vienna, Austria

Ebinuma H, Sugita K, Matsunaga E et al. (1997) Selection of marker-free transgenic plants using the isopentenyl transferase gene. Proc Natl Acad Sci U S A 94:2117–2121

Ebinuma H, Sugita K, Matsunaga E et al. (2001) Systems for the removal of a selection marker and their combination with a positive marker. Plant Cell Rep 20:383–392

Ebinuma H, Sugita K, Matsunaga E et al. (2004) Asexual production of marker-free transgenic aspen using MAT vector systems. In: Kumar S, Fladung M (eds) Molecular genetics and breeding of forest trees. Food Products, New York, pp 309–338

Ebinuma H, Sugita K, Endo S et al. (2005) Elimination of marker genes from transgenic plants using MAT vector systems. In: Pena L (ed) Transgenic plants. Humana, New Jersey, pp 237–253

Eszterhas S-K, Bouhassira E-E, Martin D-I et al. (2002) Transcriptional interference by independently regulated genes occurs in any relative arrangement of the genes and is influenced by chromosomal integration position. Mol Cell Biol 22(2):469–479

Feng Y-Q, Seibler J, Alami R et al. (1999) Site-specific chromosomal integration in mammalian cells: highly efficient Cre recombinase-mediated cassette exchange. J Mol Biol 292:779–785

Fukushige S, Sauer B (1992) Genomic targeting with a positive-selection *lox* integration vector allows highly reproducible gene expression in mammalian cells. Proc Natl Acad Sci U S A 89:7905–7909

Feng Y-Q, Lorincz M-C, Fiering S et al. (2001) Position effects are influenced by the orientation of a transgene with respect to flanking chromatin. Mol Cell Biol 21(1):298–309

Kim S-I, Veena, Gelvin S-B (2007) Genome-wide analysis of *Agrobacterium* T-DNA integration sites in the *Arabidopsis* genome generated under nonselective conditions. Plant J 51:779–791

Kumar S, Fladung M (2001) Gene stability in transgenic aspen (*Populus*). II. Molecular characterization of variable expression of transgene in wild and hybrid aspen. Planta 213:731–740

Kohli A, Twyman R-M, Abranches W et al. (2003) Transgene integration, organization and interaction in plants. Plant Mol Biol 52:247–258

Louwerse J-D, Van Lier M-C-M, Van Der Steen et al. (2007) Stable recombinase-mediated cassette exchange in *Arabidopsis* using *Agrobacterium tumefaciens*. Plant Physiol 145:1282–1293

Lyznik L-A, Gordon-Kamm W-J, Tao Y (2003) Site-specific recombination for genetic engineering in plants. Plant Cell Rep 21:925–932

Maqbool S-B, Christou P (1999) Multiple traits of agronomic importance in transgenic indica rice plants analysis of transgene integration patterns, expression levels and stability. Mol Breed 5:471–480

Matsuzaki H, Nakajima R, Nishiyama J et al. (1990) Chromosome engineering in *Saccharomyces cerevisiae* by using a site-specific recombination system of a yeast plasmid. J Bact 172:610–618

Matzke M-A, Matzke A-J-M (1991) Differential inactivation and methylation of a transgene in plants by two suppressor loci containing homologous sequences. Plant Mol Biol 16:821–830

Matzke M-A, Mette M-F (2000) Transgene silencing by the host genome defense: implications for the evolution of epigenetic control mechanism in plants and vertebrates. Plant Mol Biol 43:401–415

Meyer P (2000) Transcriptional transgene silencing and chromatin component. Plant Mol Biol 43:221–234

Nanto K, Ebinuma H (2007) Expression of a transgene exchanged by the recombinase-mediated cassette exchange (RMCE) method in plants. International Conference - Plant Tansformation Technologies (Abstract: pp 71) Vienna, Austria

Nanto K, Ebinuma H (2008, on line: 2007) Marker-free site-specific integration plants. Transgenic Res 17:337–344

Nanto K, Yamada-Watanabe K, Ebinuma H (2005) *Agrobacterium*-mediated RMCE approach for gene replacement. Plant Biotechnol J 3:203–214

Nanto K, Sato K, Katayama Y, Ebinuma H (2009) Expression of a transgene exchanged by the recombinase-mediated cassette exchange (RMCE) method in plants. Plant Cell Rep 28:777–785

O'Gorman S, Fox D-T, Wahl G-M (1991) Recombinase-mediated gene activation and site-specific integration in mammalian cells. Science 251:1351–1355

Onouchi H, Yokoi K, Machida C et al. (1991) Operation of an efficient site-specific recombination system of *Zygosaccharomyces rouxii* in tobacco cells. Nucleic Acids Res 19:6373–6378

Ow D-W (2002) Recombinase-directed plant transformation for the post-genomic era. Plant Mol Biol 48:183–200

Schlaman H-R-M, Hooykaas P-J-J (1997) Effectiveness of the bacterial gene *codA* encoding cytosine deaminase as a negative selectable marker in *Agrobacterium*-mediated plant transformation. Plant J 11:1377–1385

Srivastava V, Ow D-W (2001) Biolistic mediated site-specific integration in rice. Mol Breed 8:345–350

Srivastava V, Ow D-W (2004) Marker-free site-specific gene integration in plants. Trends Biotech 22:627–629

Srivastava V, Ariza-Nieto M, Wilson A-J (2004) Cre-mediated site-specific gene integration for consistent transgene expression in rice. Plant Biotechnol J 2(2):169–179

Sugita K, Matsunaga E, Ebinuma H (1999) Effective selection system for generating marker-free transgenic plants independent of sexual crossing. Plant Cell Rep 18:941–947

Tzfira T, Frankman L-R, Vaidya M et al. (2003) Site-specific integration of *Agrobacterium tumefaciens* T-DNA via doublestranded intermediates. Plant Physiol 133:1011–1023

Vain P, James V-A, Worland B et al. (2002) Transgene behaviour across two generations in a large random population of transgenic rice plants produced by particle bombardment. Theor Appl Genet 105:878–889

Vergunst A-C, Hooykaas P-J (1998) Cre/lox-mediated site-specific integration of *Agrobacterium* T-DNA in *Arabidopsis thaliana* by transient expression of cre. Plant Mol Biol 38:393–406

Vergunst A-C, Jansen L-E, Hooykaas P-J (1998) Site-specific integration of *Agrobacterium* T-DNA in *Arabidopsis thaliana* mediated by Cre recombinase. Nucleic Acids Res 26:2729–2734

Chapter 23
Promoter Trapping in Plants Using T-DNA Mutagenesis

R. Srinivasan and Dipnarayan Saha

Abstract Promoter is a DNA fragment with crucial *cis*-acting signature sequences governing the transcription of an adjoining gene. The promoter elements govern the level of expression of the associated gene, determine the responsiveness of the gene to physical and/or chemical stimuli and decide whether the associated gene would constitutively express or would exhibit tissue or developmental stage specific expression. Depending upon the signal sequences present in the promoter element the expression of the associated gene would respond to one or to a combination of more than one such signal. Promoter element is thus the key component for regulated expression of introduced genes in transgenic plants. A range of promoter elements is being identified and exploited for need-based tailoring of transgenes both for basic and applied purposes. A variety of approaches are employed to isolate promoter elements from plants, amongst them transferred DNA (T-DNA) insertion mutagenesis has been extensively used in identification and cloning of genes, promoters, enhancers and other regulatory sequences. In this chapter, we discuss the T-DNA based approach for identification and isolation of promoter elements from plants. We also highlight the impact of T-DNA tagging strategy in functional genomics studies and discuss some recent developments in the field of cloning, characterization and identification of promoter elements in *Arabidopsis thaliana*. Finally, we summarize some of the commonly employed approaches for characterization and analysis of plant promoters.

23.1 Introduction

Promoter is an indispensable part of a functional gene. A promoter consists of a core sequence and several *cis*-regulatory sequences responsible for binding of proteins called transcription factors (TFs). The expression of a gene is determined by the presence of

R. Srinivasan (✉) and D. Saha
National Research Centre on Plant Biotechnology, Indian Agricultural Research Institute Campus, New Delhi, 110012, India
e-mail: srinivasan53@gmail.com

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_23, © Springer Science+Business Media B.V. 2010

specific signature sequences in the promoter region and their complex but coordinated interactions with the TFs (Li et al. 2002). Based on the nature of gene expression, promoters have been classified as constitutive, tissue specific, temporal/stage specific or inducible promoters. Constitutive promoters are responsible for regulating gene expression throughout a plant life in all the cell types, whereas temporal and tissue specific promoters regulate gene expression in certain developmental stages or in certain cell types. The inducible promoters are activated only upon sensing some physical, chemical or environmental cues (Datla et al. 1997; Tyagi 2001). Some of these promoter elements are able to retain their functionality across the species, genera and kingdom e.g., the *35S cauliflower mosaic virus* (*35S CaMV*) promoter is constitutively expressed in several plants and is probably the most popularly used regulatory element in transgenic experimentations (Yoshida and Shinmyo 2000).

The history of plant promoter identification and elucidation of plant genes dates back to late 1970s to early 1980s with the landmark discovery of *Agrobacterium* mediated transfer of genes and their expression in plants (Kuhlemeier et al. 1987; Ream and Gordon 1982). The fine mapping of the T-DNA region led to the identification and characterization of promoter regions of *octopine synthase* (*ocs*) and *nopaline synthase* (*nos*) genes, which have since been widely exploited in plant genetic engineering (Bevan et al. 1983; Herrera-Estrella et al. 1983; Fraley et al. 1983). The conventional approach for identification of promoter is to clone the adjoining DNA sequence from a genomic DNA library using the transcribed part of the gene as a probe. This approach has led to the isolation of several promoter elements e.g. promoters specific for anther tissues (Albani et al. 1992; Koltunow et al. 1990; Beals and Goldberg 1997). In recent years, the advancement in the field of functional genomics has led to a variety of techniques for promoter isolation in plants. Some of them involve prior information about transcripts either through differential display technique (Reuber and Ausubel 1995) or microarray-based expression profiling (Schena et al. 1995; Richmond and Sommerville 2000). The transcript information is utilized for cloning of the gene and its regulatory regions (Digeon et al. 1999; Mishra et al. 2002; Ukai et al. 2002).

One of the major bottlenecks in conventional approach of promoter isolation is the difficulty in the identification of promoter element of an individual gene belonging to a gene family. Even the recently developed techniques like microarray and real time PCR-based approach of transcript identification is technically demanding and limited by the sensitivity of the method to detect genes that are expressed at a very low level and/or in a transient manner (Springer 2000). In this context, a complementary approach of random insertion of reporter genes either through T-DNA or transposons(Tn) as mutagen is exploited for plant gene promoter isolation (Azpiroz-Leehan and Feldmann 1997; Krysan et al. 1999; Resminath et al. 2005a; Parinov and Sundaresan 2000). In principle, the reporter gene present in the T-DNA or Tn gets inserted in the vicinity of a regulatory sequence and under its influence gets transcribed. The reporter gene in a T-DNA or transposon in the process literally traps or tags an element conferring the expression behavior. Difference in the ‘trapping’ constructs determines the type of DNA element identified. A variety of vectors have been designed to trap and identify genes, exons, or the regulatory elements like promoters and enhancers (Springer 2000).

The salient feature of T-DNA based promoter trapping and its functional utility in plant biology has been addressed in the present article. Besides providing the recent information on promoter and enhancer cloning in plant using T-DNA-based promoter trapping and enhancer trapping strategies, we also discuss the most commonly employed approaches for characterization of plant promoters.

In recent years, the genome sequence information in several organisms has effectively complemented genomic insertion tools and greatly facilitated identification and isolation of genes and promoters. The major advantage of these tools is that they allow easy identification of genes and promoters and assist precise functional annotation of the tagged element if the genomic sequence information is available. In plants, the T-DNA of *Agrobacterium* and transposon insertions have significantly contributed to isolation of genes and promoters through gene and promoter trap vectors, respectively (Fobert et al. 1991; Klimyuk et al. 1995; Resminath et al. 2005b; Sundaresan et al. 1995; Topping et al. 1991). In the T-DNA- based promoter-trapping approach the constructs consist of a promoterless reporter gene for easy monitoring of the activity of the tagged chromosomal locus. Table 23.1 summarizes the features of promoter, enhancer and gene trap vectors.

The event of T-DNA integration downstream to a native gene promoter results in either transcriptional or translational gene fusions, leading to the expression of the reporter gene (Fig. 23.1). The T-DNA insertion within the exonic sequence leads to transcriptional fusion between the reporter gene and the mRNA coding gene, while the translational gene fusion takes place when the reporter gene lacking an ATG start codon is integrated in frame with the coding sequence of a gene

Table 23.1 Comparison between promoter trap, enhancer trap and gene traps

Feature	Promoter trap	Enhancer trap	Gene trap
Vector strategy	Consists of reporter gene without any promoter element	Reporter gene fused to a minimal promoter	Consists of promoterless reporter gene and a splice acceptor site
Expression condition	Insertion in correct orientation and in a transcriptional unit or downstream to a native gene promoter	Insertion in any orientation and in any place within the range of an enhancer element activity	Insertion in correct orientation and within a gene nearby a splice site donor
Probability of finding reporter gene expression	Low	High	Low
Probability of gene knockouts	High	Low	High
Isolation of responsible *cis*-acting element	Easy	Difficult; requires careful and detailed experimentations	Easy

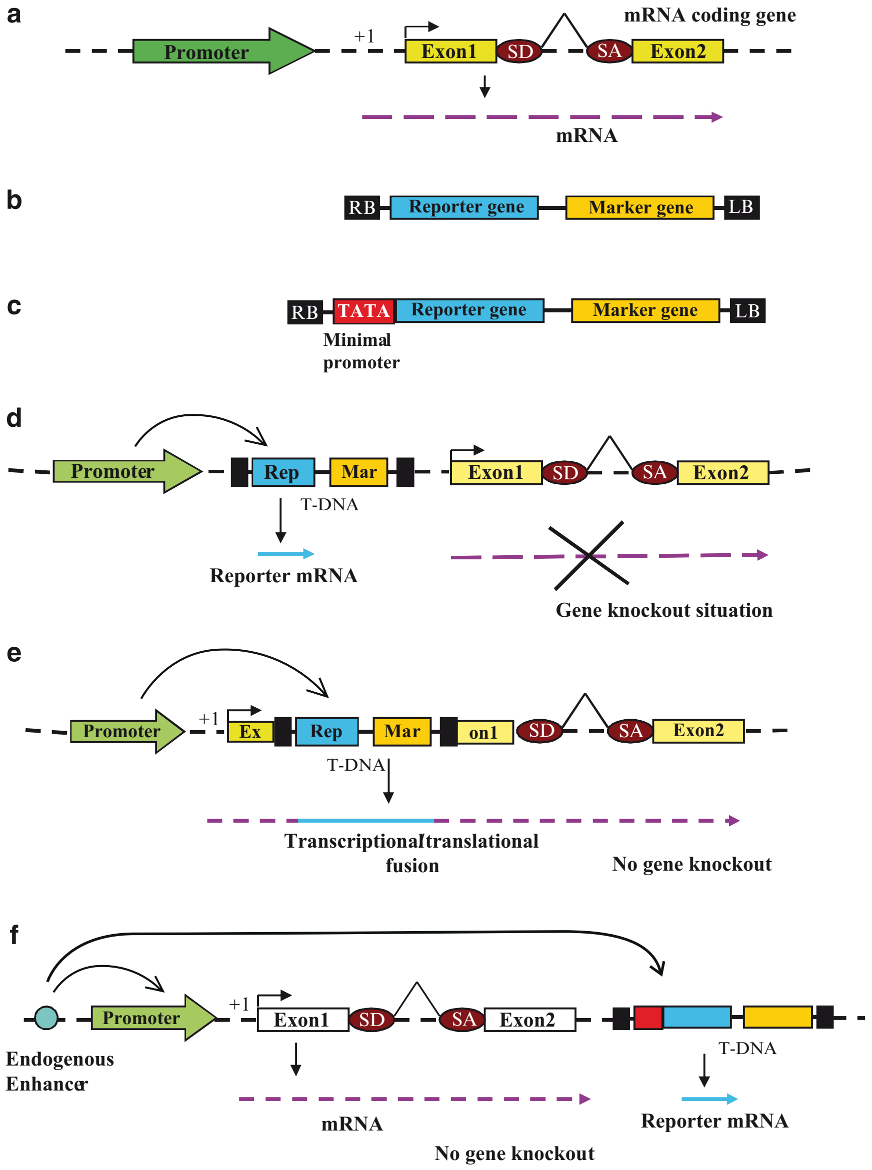


Fig. 23.1 Schematic *representation* of a typical *T-DNA of* promoter and enhancer trap constructs. (**a**) Simplified structure of a typical mRNA coding gene. SD and SA are splice donor and splice acceptor site involved in intron splicing. The +1 represents the TSS. (**b**) A schematic diagram of T-DNA based promoter trapping construct with promoterless reporter gene. RB and LB are right border and left border of T-DNA. (**c**) Simplified diagram representing T-DNA based enhancer trapping construct. The TATA box represents the minimal promoter sequence containing only TATA box fused with the reporter gene. (**d**) A gene knockout situation resulting from the T-DNA based promoter trap insertion in the adjacent 5′ upstream sequence of a nearby gene leading to the disruption of transcript production of the gene. The adjacent promoter sequence however activates the synthesis of reporter gene transcript. (**e**) A situation resulting from T-DNA insertion in the initial sequence of the exon giving rise to transcriptional/translational fusion of reporter and gene product. The situation *need* not result into complete gene knockout. (**f**) An enhancer trap construct may be influenced by an endogenous enhancer element situated far apart from the T-DNA insertion site

(Koncz et al. 1989). Besides getting transcribed and activated, the promoter trap elements may also result into knockout effect of the gene to produce a mutant phenotype when inserted into the structural part of a gene.

The first generation promoter trapping constructs used in plants consisted of promoterless selectable marker gene placed nearby the borders of the T-DNA region to capture a regulatory sequence adjacent to the T-DNA junction (Andre et al. 1986; Herman et al. 1990; Koncz et al. 1989). One of the limitations with such approach was that the transgenics were selected only when the marker gene was placed adjacent to an active regulatory DNA sequence. As a consequence trapped promoter activity was detected only during the time of selection pressure and promoters active at early developmental events or not expressed during callus phase or specifically expressed in specialized cells/tissues or responding to specific signals etc. could not be detected (Koncz et al. 1989). The above limitation was addressed subsequently by employing an additional marker gene driven by a constitutive promoter to allow the selection of transformants, while the promoterless marker gene placed adjacent to the border sequences facilitated tagging of an active promoter (Koncz et al. 1989). However, the major drawback of using resistance marker gene for promoter trapping is that the antibiotic resistance does not allow us to distinguish and identify the mutated phenotype in plants heterozygous for T-DNA insertions. The present generation promoter trap vectors employ visible and scorable reporter genes like *uidA* gene coding for β-glucuronidase (GUS)(Fobert et al. 1994; Kertbundit et al. 1991; Lindsey et al. 1993; Plesch et al. 2000; Topping et al. 1991; Wei et al. 1997), green fluorescent protein (*gfp*) (Haseloff et al. 1997; Stewart 2001; Yang et al. 1996) and *luciferase* (Alvarado et al. 2004; Mudge and Birch 1998; Remy et al. 2005; Yamamoto et al. 2003) which allow easy histological screening of the trapped promoter activity even in plants heterozygous for T-DNA insertions (Walden 2002).

The enhancer trap vectors are similar to the promoter traps except the enhancer trap consists of minimal promoter sequence. The minimal promoter sequence by itself is not sufficient to drive reporter gene expression but under the influence of *cis*-regulatory DNA sequences like enhancers, can activate the reporter gene (Fig. 23.1). Enhancer trapping approach employs either T-DNA (Campisi et al. 1999) or transposon systems like *Dissociator* (*Ds*) elements (Fedoroff and Smith 1993; Parinov et al. 1999; Sundaresan et al. 1995; Yang et al. 2005). The Tn-based enhancer traps have been exploited in plant species with well-characterized Tn elements. Alternatively, the Tn elements from heterologous sources are cloned into T-DNA and introduced into plants to allow random transposition in the plant genome (Osborne and Baker 1995).

23.2 T-DNA-Based Promoter Trapping: Advantages and Limitations

The advantages of the present day T-DNA-based promoter tagging in plants include easy identification of promoters even those, which direct the transcription only under certain conditions and are normally difficult to identify through conventional approaches (Table 23.2). For example, promoters (i)that are active at different plant

Table 23.2 Advantages and limitations of T-DNA based promoter trapping methods

Advantages
Rapid and simple identification of promoters. Identification is based on reporter gene expression pattern
In case of gene families and closely related genes, specifically identifies and reveals the expression pattern of an individual member
Can trap promoters that are expressed even in a single or few cells within a given tissue
Transiently and differentially expressed genes can be detected efficiently
Does not require prior knowledge of the expression of a gene
Limitations
Randomness of insertion
T-DNA insertion may result in chromosomal rearrangements

developmental stages (Campisi et al. 1999); (ii) in response to different environmental stimuli (Mandal et al. 1995); (iii) induced upon nematode (Barthels et al. 1997; Puzio et al. 2000) and (iv) pathogen infection (Custers et al. 2002) etc., have been isolated through T-DNA promoter trapping. The T-DNA promoter traps can also lead to the identification and isolation of several cryptic promoter elements in plants, which otherwise are silent and apparently not associated with any protein or RNA coding gene (Foster et al. 1999; Mollier et al. 2000; Okresz et al. 1998; Sivanandan et al. 2005).

One of the major limitations of the T-DNA-based promoter isolation process is the non-specific and random nature of T-DNA integration (Table 23.2). At this juncture it is not been possible to efficiently target the insertion to a specific genomic region. Whether the nature of T-DNA insertion is random or preferred for the transcribed genic regions is still an ongoing debate (Francis and Spiker 2005). The efficiency of promoter traps are limited by the fact that the activity of the reporter gene in a promoter trap is detected only when the reporter gene is integrated in a correct orientation and downstream to an active native gene promoter. The altered phenotype of a plant because of T-DNA insertion provides useful clues towards the function of the gene. However, in some cases the pattern of reporter gene expression provides primary information about the specificity and function of the promoter and the corresponding gene even in the absence of a mutant phenotype (Prasad et al. 2005). Another major problem, commonly encountered during the T-DNA integration process is the occurrence of extensive chromosomal DNA rearrangements like deletions, duplications, inversions and reciprocal translocations (Nacry et al. 1998; Takano et al. 1997). These chromosomal rearrangements render the subsequent process of identification and characterization of promoter elements very difficult.

23.3 T-DNA Promoter Trapping

The requirements and the process of T-DNA mediated promoter trapping in plants is outlined in Fig. 23.2: (i) an appropriate trapping vector consisting of a promoterless reporter gene and a selectable marker gene for the selection of transformants.

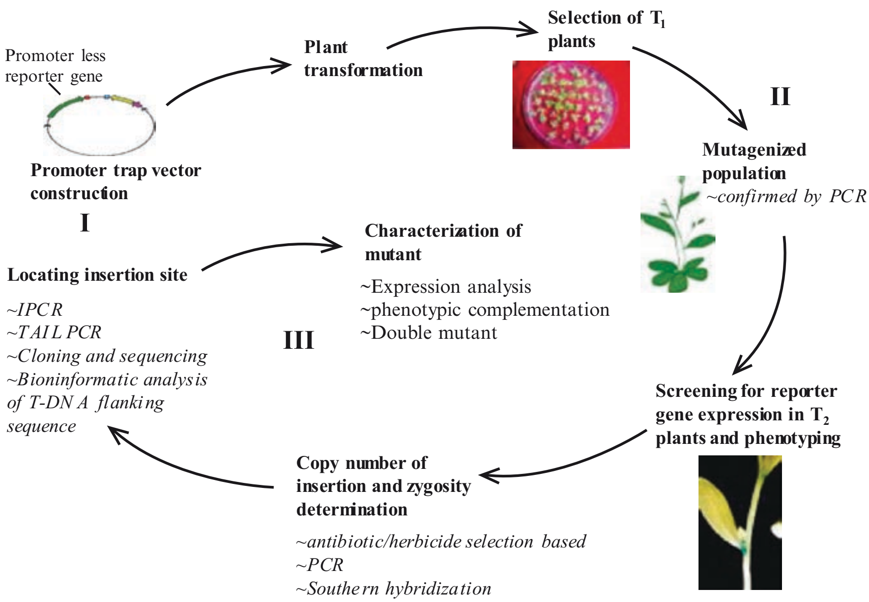


Fig. 23.2 An overview of steps involved in generating mutagenized population of T-DNA based promoter trapped lines in *Arabidopsis thaliana* towards isolation of promoter elements**.** The overall process involves three basic steps; step I involves construction of suitable trapping vector, step II involves transformation, screening of transformants and generation of mutant population. The step III involves characterization of mutants and localization of T-DNA insertions in the plant chromosomal DNA

(ii) A suitable method for introduction of the T-DNA into plant genome either through *Agrobacterium* co-cultivation of explants or in planta transformations followed by screening of the transformants to generate T-DNA tagged mutant population, (iii) identification of a suitable mutant based on expression pattern of the reporter gene, and (iv) Identification and characterization of the flanking sequences of the plant genome.

23.3.1 Arabidopsis Thaliana

The availability of the complete genome sequence and the ease of generating large population of transformed plants make *A. thaliana* an excellent plant system for isolation of genes and regulatory elements through T-DNA gene traps (Walden 2002). More than 320,000 insertional mutations are available in the *A. thaliana Columbia* genotype, which are potential sources for identification of genes and regulatory elements (Bevan and Walsh 2005). Table 23.3 lists some of the promoters isolated through T-DNA mediated promoter trapping in *Arabidopsis*.

The T-DNA based promoter trapping has been used in isolating promoter elements conferring gene expression in specific cell types and tissues. For example, regulatory elements were identified specifically for seed tissues like endosperm, embryo and other seed tissues in *Arabidopsis* (Casson et al. 2002; Stangeland et al. 2003; Topping et al. 1994; Topping and Lindsey 1997). Other significant examples of tissue specific promoter elements isolated through T-DNA promoter trapping in *Arabidopsis* includes phloem tissue-specific promoter (Kertbundit et al. 1991), *caffeic acid O-methyltransferase 1* (*AtOMT1*) gene promoter expressing in root vascular tissues (Goujon et al. 2003) and *elF-4A1* promoter conferring gene expression in growing tissues and young leaves (De Greve et al. 2001).

In *Arabidopsis*, the *POLARIS* (*PLS*) and *EXORDIUM* (*EXO*) gene, which express specifically in the basal embryonic regions and seedling root tip (Casson et al. 2002) and in dividing cells (Farrar et al. 2003), respectively were identified and cloned from a promoter-trapped line. Other examples include a *LATERAL ROOT PRIMORDIUM 1* (*LRP1*) gene showing specific expression in lateral roots of *Arabidopsis* (Smith and Federoff 1995) and the tapetum and vascular tissues specific nucleic acid helicase gene encoding HVT1 (Helicase in Vascular tissue and Tapetum) protein in *Arabidopsis* (Wei et al. 1997).

Using T-DNA promoter trapping tool in *Arabidopsis*, our laboratory has generated a few mutants exhibiting reporter gene GUS in different tissues (Resminath et al. 2005a). Characterization of these mutants has led to identification of promoter elements conferring gene expression in specific tissues like lateral organ junctions (Prasad et al. 2005), roots (Sivanandan et al. 2005), anthers (Thakare et al. 2006) and trichomes. The regulatory elements present in the upstream region of the *LOJ* gene conferring characteristic tissue specificity were characterized (Prasad et al. 2005; Saha et al. 2007).

Other than tissue-specific promoters, the promoter trapping in *Arabidopsis* has led to the isolation of conditionally inducible promoters. For example, the promoters induced by pathogen (Custers et al. 2002), nematode feeding (Barthels et al. 1997; Favery et al. 1998; Goddijn et al. 1993; Puzio et al. 1999, 2000) and responsive to different environmental stimuli like low temperature and abscisic acid (Mandal et al. 1995) etc., have been isolated. T-DNA promoter trapping serves as a convenient option to identify pseudo-promoters or cryptic promoters (Lindsey et al. 1993). Several cryptic regulatory elements in *Arabidopsis* have been trapped by T-DNA and identified on the basis of reporter gene expression pattern. These include promoter elements specific for meristematic tissues like calli and root tips (Okresz et al. 1998) guard cells (Plesch et al. 2000) and roots (Mollier et al. 2000; Sivananadan et al. 2005).

23.3.2 Tobacco

Prior to the popularity of *Arabidopsis* as a model plant, tobacco served as an experimental plant in which the promoter trapping tool was widely employed for

Table 23.3 List of some important plant promoters cloned by promoter trapping approach

Plant species	Isolated promoter	Expression in tissues	Reporter gene	References
A. thaliana	AtEM101,201	Zygotic embryo	uidA (*GUS*)	Topping et al. (1991, 1994)
N. tabacum	*Cryptic*	Seed coat	uidA (*GUS*)	Fobert et al. (1994)
A. thaliana	AtVT1	Vascular and tapetum	uidA (*GUS*)	Wei et al. (1997)
N. tabacum	tCUP	Constitutive	uidA (*GUS*)	Foster et al. (1999)
A. thaliana	Pyk20	Nematode feeding organ/tissues	uidA (*GUS*)	Puzio et al. (2000)
A. thaliana	*Cryptic*	Root	uidA (*GUS*)	Mollier et al. (2000)
A. thaliana	*Cryptic*	Guard cell	uidA (*GUS*)	Plesch et al. (2000)
L. japonicus	Lj Cbp1	Roots	uidA (*GUS*)	Webb et al. (2000)
A thaliana	elF-4A1	Developing tissues and young leaves	uidA (*GUS*)	DeGreve et al. (2001)
A. thaliana	488-*promoter*	Pathogen induced	uidA (*GUS*)	Custers et al. (2002)
A. thaliana	EXORDIUM *promoter*	Embryo and meristematic cells	uidA (*GUS*)	Farrar et al. (2003)
A. thaliana	*Cryptic*	Root	uidA (*GUS*)	Sivanandan et al. (2005)
A. thaliana	LOJ *promoter*	Lateral organ junctions	uidA (*GUS*)	Prasad et al. (2005)
A. thaliana	Anth85	Anther	uidA (*GUS*)	Thakare et al. (2006)
L. japonicus	LjENOD40–2	Root and nodule	uidA (*GUS*)	Buzas et al. (2005)

activation and identification of gene and its regulatory sequences (Herman et al. 1990; Fobert et al. 1991; Lindsey et al. 1993). The ease of regeneration and high frequency transformation makes tobacco plant a favourite for promoter trapping (Topping et al. 1991; Topping and Lindsey 1995; Walden 2002). T-DNA promoter trapping in tobacco has led to isolation of prmoter specific floral organs (Lindsey et al. 1993), constitutive cryptic promoter like *tCUP* (Fobert et al. 1999) and a cryptic promoter responsible for driving reporter gene expression in tobacco seed coats in a spatial and temporal manner (Foster et al. 1994).

23.3.3 Other Plants

The significant achievements of T-DNA promoter tagging in model plants like *Arabidopsis* and tobacco has led to its extension to other plants as well. Promoter trapping in the legume plant *Lotus japonicus*, which is regarded as a model plant for genetic analysis of nodulation studies (Handberg and Stougaard 1992) has led to the identification of regulatory elements specific for roots and nodules (Martirani et al. 1999; Buzas et al. 2005). Using the T-DNA promoter trap method, Bade et al. (2003) have generated a collection of mutants in *Brassica napus*. Mutants exhibiting constitutive and callus specific reporter gene expression patterns have been characterized. In tomato and potato, the promoter trapped mutant lines exhibited reporter gene expression in flowers, fruits and seedlings (Meissner et al. 2000) and in vegetative and floral organs, respectively (Lindsey et al. 1993). In the case of poplar, the promoter trapping approach has resulted in the identification of genes expressed specifically in the vascular cambium tissues and their respective regulatory elements (Johansson et al. 2003). Although the T-DNA tagging mediated promoter isolation is increasingly being exploited in model plant *Arabidopsis* and is being also used in other systems the real challenge is to apply this technique in other plants with higher efficiency.

23.4 T-DNA Based Enhancer Trapping in Plants

A specialized entrapment construct comprising of a minimal promoter with relatively weak activity fused to a reporter gene within the T-DNA border region (Campisi et al. 1999; Goldsbrough and Bevan 1991; Klimyuk et al. 1995; Parinov et al. 1999; Sundaresan et al. 1995; Topping et al. 1991) facilitates tagging of enhancer elements. The interesting advantage of enhancer trap lines is that they specify cell and tissue specific markers especially for the plant organs with complex structures like embryo, flower and ovule (Campisi et al. 1999). Based on the tool to introduce the vector in the plant genome, the enhancer traps can be classified as T-DNA based enhancer traps and the transposon (Tn) based enhancer traps. The transposon based traps consists of transposable elements like *Dissociator* (*DS*) ele-

ments and *Activator* (*Ac*) elements either in a single cassette or in two separate cassettes (Springer et al. 2000). In two-component enhancer trap system, the *Ds* element carrying the minimal promoter linked reporter gene is activated upon genetic crossing with a plant carrying *Ac* element (Klimyuk et al. 1995; Parinov et al. 1999). The random transposition of *Ds* element throughout the plant genome allows tagging of functional enhancer elements through detection of reporter gene expression (Fig. 23.3).

The two component Tn-element-based enhancer trapping in *Arabidopsis* resulted into generation of large number of enhancer detector lines, which led to the identification of an egg apparatus specific enhancer element, *EASE* (Yang et al. 2005). The identification of *EASE* was speculated to be a promising tool for manipulating embryo specific traits towards exploitation of apomixes. The other instances of with Tn-based enhancer trapping involve isolation of novel genes and regulatory elements in poplar (Groover et al. 2004; Filichkin et al. 2006), and rice (Upadhyaya et al. 2002).

As a two-component system the enhancer trap vectors have been improved to employ GAL4 and UAS (Upstream Activator Sequence) along with a nondestructive reporter gene like *green fluorescent protein* (*gfp*) or *luciferase* (*luc*) (Engineer et al. 2005; Laplaze et al. 2005). The GAL4/UAS system facilitates efficient enhancer trapping and allows a quantifiable reporter gene expression in plants (Engineer et al. 2005). Using GAL4-UAS enhancer trapping system and *gfp* as reporter gene Laplaze et al. (2005) have identified lateral root specific enhancer element in the *LOB-domain protein16* gene (*LBD16*) promoter of *Arabidopsis*. Similarly a large collection of enhancer trapped mutant lines were developed in rice using GAL4/VP16-UAS as enhancer trap and *uidA* as reporter gene (Wu et al. 2003). The reporter gene expression was detected in tissues of leaves, roots and floral organs indicating variety of regulatory element tagged.

T-DNA based enhancer traps have been employed to generate large populations of mutants in tobacco (Topping et al. 1991), potato (Goldsbrough and Bevan 1991) and *Arabidopsis* (Campisi et al. 1999). From the *Arabidopsis* enhancer trap lines, an enhancer element associated with D-class cyclin gene exhibiting vegetative and floral primordia specific expression has been isolated (Swaminathan et al. 2000). T-DNA enhancer trapping have led to the identification of plant lines exhibiting reporter gene expression associated with circadian cycle in *Arabidopsis* (Michael and McClung 2003) and specific to embryogenic calli in carrot (Ko and Kamada 2002), respectively.

23.5 Promoter Cloning Strategies in Plants

Once a promoter is tagged with T-DNA, the T-DNA sequence becomes a potential target for cloning of the flanking sequences of plant genome. The flanking sequences are more likely to harbour the promoter elements. Some of the important plant promoters isolated and cloned by insertional mutagenesis approaches are listed in Table 23.3.

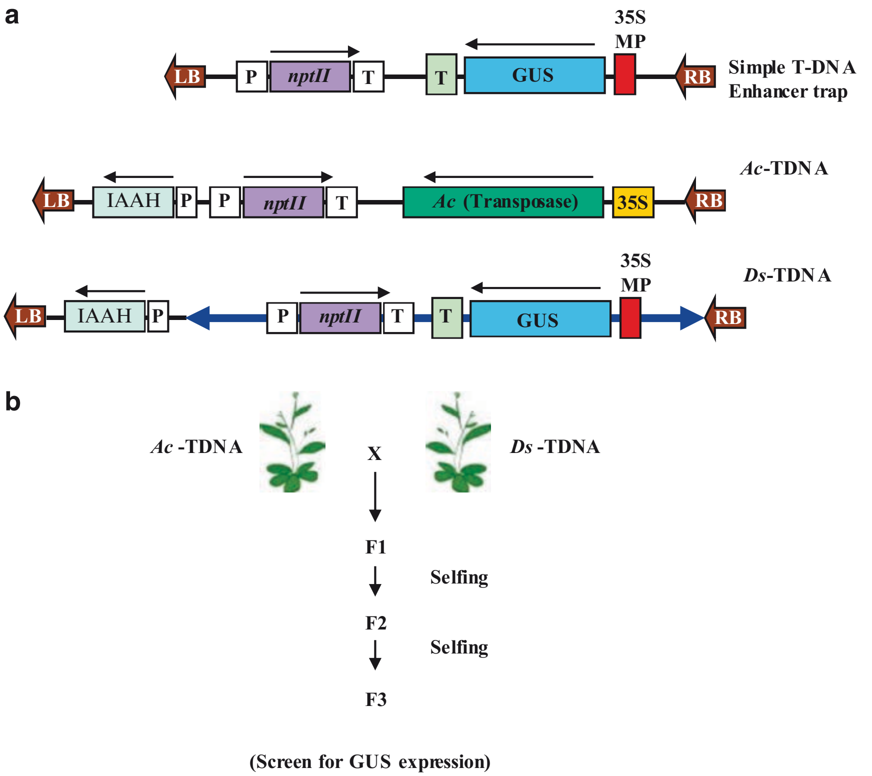


Fig. 23.3 Enhancer trapping strategies in plants. (**a**) Comparative features of enhancer trap vectors. A simple T-DNA based enhancer traps consists of only reporter gene driven by a minimal promoter sequence. The two component enhancer trap involves Ac and Ds in separate constructs. P: promoter, T: terminator, 35S MP: truncated 35S CaMV promoter sequence as minimal promoter containing only TATA sequence. Black arrows indicate directions of the transcription. The transposable element Ds is represented by thick blue double headed arrow. (**b**) Mobilization scheme of enhancer trap by crossing the individual plants harboring Ac element and Ds element separately. The Ac element helps in random transposition of the Ds element in the genome. The selfed plants serves as enhancer trap mutant lines.

Several strategies are available to isolate and clone the T-DNA flanking genomic DNA sequences (Fig. 23.4). (i) Library screening of the mutant plant genomic DNA is one of the approaches for promoter cloning. The DNA fragments flanking the T-DNA are identified from the library and used as a probe to isolate the wild type genomic sequence. The above approach is cumbersome but is particularly suitable when tandemly repeated copies of T-DNA are present at the same location (Marks and Feldman 1989). (ii) Plasmid rescue is another approach that is employed in case the T-DNA construct carries a bacterial antibiotic resistance gene and a bacterial origin of replication (*ori*). The genomic DNA of mutant plant is subjected to complete digestion by a suitable restriction enzyme followed by ligation to circularize all the fragments and transform them into *E. coli* host. Only the circular

DNA fragments carrying the *ori* sequence and the resistance gene in the T-DNA can survive in *E. coli* on antibiotic selection medium. The plasmids isolated from the *E. coli* are analyzed for the presence of T-DNA and the flanking plant DNA sequences (Dilkes and Feldmann 1998; Yanofsky et al. 1990). (iii) Compared to the above two methods, a more reliable and easy approach for isolating plant T-DNA flanking sequences is based on PCR based technique known as inverse PCR (IPCR). The IPCR strategy principally involves enzymatic cleavage within the T-DNA region followed by ligation of the fragments to facilitate self-circularization. A set of nested primers derived from the T-DNA border regions are used to amplify the flanking DNA(Ochman et al. 1988; Resminath et al. 2005b; Wei et al. 1997); (iv) the thermal asymmetric interlaced PCR (TAIL-PCR) is perhaps the most sensitive technique among the all for recovering the T-DNA flanking genomic sequences. The technique makes use of three nested T-DNA-specific primers in one end and a short arbitrary degenerate (AD) primer in the other end. Three different PCR reactions are performed with these primer sets. The primary PCR reaction involves different primer annealing temperatures and low and high stringent cycles to facilitate annealing of arbitrary and specific primers, respectively. This step results into both specific as well as nonspecific amplification of products. In the next two steps of PCR reactions the non-specific products are eliminated amplifying predominantly the T-DNA flanking genomic DNA (Liu et al. 1995). (v) The other PCR-based method of promoter cloning includes enzymatic blunting of mutant genomic DNA and ligation of adapters followed by amplification of the T-DNA flanking DNA using three nested primers specific to T-DNA and adapters (Siebert et al. 1995). In al the cases, the PCR products obtained is cloned, sequenced to identify the genomic sequences flanking the T-DNA and subjected to further analysis.

23.6 Characterization of Plant Promoters

23.6.1 In Silico Analysis

In silico analysis is employed to delineate the nature of the flanking sequences. Once the putative coding sequence of the associated gene is identified, the upstream sequence is invariably subjected to in silico analysis using different promoter prediction algorithms. The commonly employed analysis include identification of potential *cis*-acting regulatory elements (CAREs), TFs and putative transcription start site (TSS) (Fickett and Hatzigeorgiou 1997; Molina and Grotewold 2005; Rombauts et al. 2003) (Fig. 23.5). These prediction tools are based on 'search by signals' in which the algorithms aim to identify regulatory regions based on sequence context in the subject DNA sequence. Some of the commonly used promoter prediction and analysis tools for identification of regulatory elements are listed in Table 23.4.

However, the in silico detection of CAREs does not necessarily mean all the elements detected on the DNA sequence are functionally relevant. Since, these

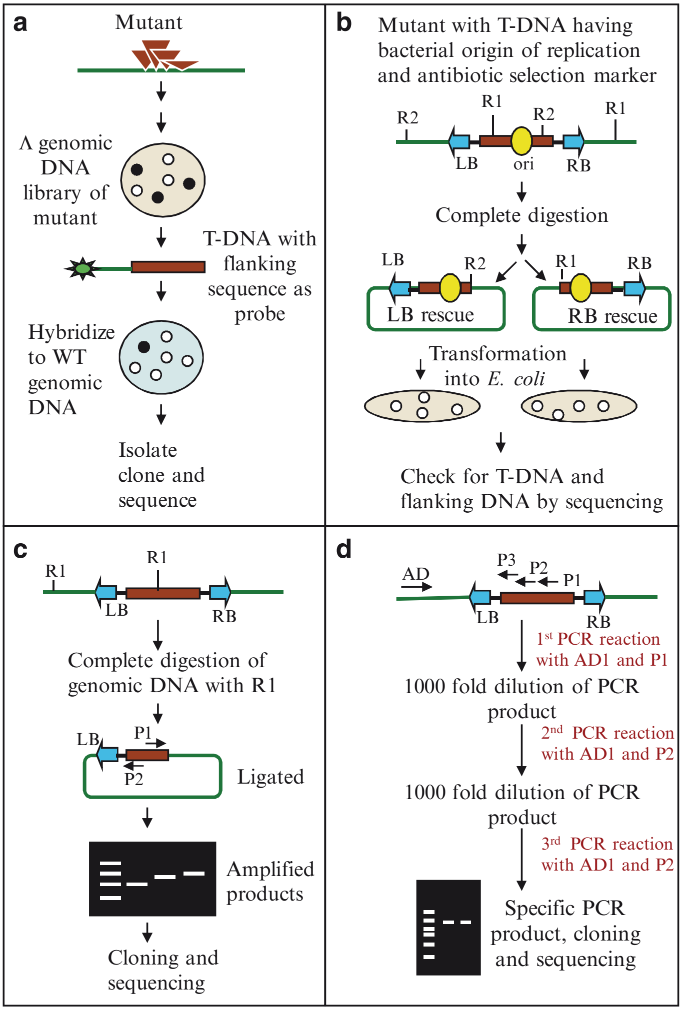


Fig. 23.4 Different approaches for cloning sequences flanking the T-DNA. (**a**) Screening of genomic library of mutant plant approach; (**b**) plasmid rescue; (**c**) inverse PCR; (**d**) TAIL-PCR. The RB and LB represents right border and left border of T-DNA; R1 and R2 are the restriction enzyme sites; P1, P2 and P3 are three T-DNA specific primers; AD represents arbitrary degenerate primer

cis-acting sequences generally consists of very short stretch of nucleotides, there is always a random chance of finding such sequences in a stretch of DNA sequence (Blanchette and Sinha 2001). One of the approaches to overcome the above limitation is to carry out a phylogenetic footprinting to find conserved regulatory elements among the functionally related promoters of diverse species or between co-expressed genes (Rombauts et al. 2003; Saha et al. 2007).

23.6.2 Determination of TSS in Plant Promoters

In general the promoter elements are known to contain the transcription start site (TSS) or +1 site, therefore locating functional TSS in the cloned DNA fragment is an essential step in the characterization of a promoter element. The techniques like primer extension, RNase protection assay, S1 nuclease analysis and 5′ rapid amplification of cDNA ends (RACE) are employed for mapping of TSS. The 5′ RACE approach being the most sensitive in finding +1 site (Carey and Smale 2000).

23.6.3 Transgenic-Based Promoter Analysis

Characterization of a promoter module and its regulatory regions requires a suitable in vivo transgenic assay system to monitor the in vivo activity level. The transgenic analysis of plant promoter includes steps like designing of vector constructs con-

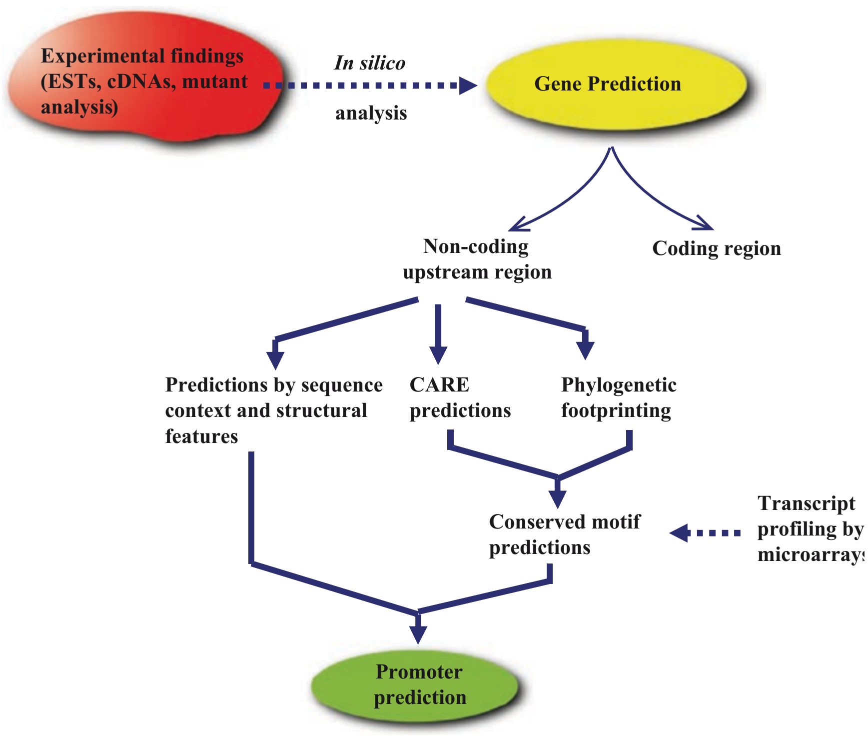


Fig. 23.5 A schematic depiction of promoter prediction through in silico tools (modified from Rombauts et al. 2003)

Table 23.4 Cis-acting regulatory element (CARE) and transcription factor binding site (TFBS) prediction programs

Prediction tools	References	URL
AtcisDB	Palaniswamy et al. (2006), Davuluri et al. (2003)	http://arabidopsis.med.ohio-state.edu/AtcisDB/index.jsp
PLACE	Higo et al. (1999)	http://www.dna.affrc.go.jp/PLACE/
PlantCARE	Lescot et al. (2002)	http://bioinformatics.psb.ugent.be/webtools/plantcare/html/
NSITE-PL (Softberry)	–	http://www.softberry.com/berry.phtml
Athena	O'Connor et al. (2005)	http://www.bioinformatics2.wsu.edu/Athena
Athamap	Galuschka et al. (2007), Steffens et al. (2004)	http://www.athamap.de/
MatInspector	Cartharius et al. (2005)	http://www.genomatix.de/products/MatInspector/index.html
TESS	Schug and Overton (1997)	http://www.cbil.upenn.edu/cgi-bin/tess/tess
JASPAR	Sandelin et al. (2004)	http://jaspar.cgb.ki.se/cgi-bin/jaspar_db.pl
TRANSFAC	Matys et al. (2003); Wingender et al. (2000)	http://www.gene-regulation.com/pub/databases.html#transfac
DATF	Guo et al. (2005)	http://datf.cbi.pku.edu.cn/index.php
Tools for phylogenetic analysis of regulatory elements		
Credo	–	http://mips.gsf.de/proj/regulomips/credo.htm
MotifSampler	Thijs et al. (2002)	http://www.esat.kuleuven.ac.be/~thijs/Work/MotifSampler.html
FootPrinter	Blanchette and Tompa (2003)	http://bio.cs.washington.edu/software.html

sisting of subject promoter fragments, a suitable method for introduction of tailor-made constructs in plant system and generation of number of transgenics and an easy but reliable assay method to monitor the pattern and level of activity (Fig. 23.6) (Pereira 2000).

Designing of vector constructs usually involves a promoterless reporter gene and a putative promoter fragment (Fig. 23.7). Based on the nature and strategy of promoter analysis, constructs could be designed for either (i) promoter deletion, (ii) linker scanning (iii) base substitution mutagenesis or (iv) identification of enhancer element in the DNA fragment. The effect of the various promoter constructs on the reporter gene expression is monitored in a plant system.

All the constructs generated are individually fused with a reporter gene and introduced into plants to monitor the reporter gene expression. The constructs could be introduced into a plant to obtain stable transformants and the expression pattern of the reporter gene analyzed in the subsequent generations, preferably in a model plant system like *Arabidopsis*, tobacco or rice.. Another approach is to introduce the recombinant constructs into plant systems for transient gene expression (Yang et al. 2000; Janssen and Gardner 1990). Although the transient expression systems provide a rapid and flexible platform for the functional analysis of the promoter, the inherent limitation of the transient assay is that the construct introduced by the process persists in an artificial configuration and in uneven copy number, leading to an inactivation or aberrant function of the introduced promoter (Kapila et al. 1997). Moreover, the transient expression does not reflect the true tissue specificity obtained *in vivo* and exists only for a shorter period of time up to 48–72 h, making it difficult to monitor promoter activity over a longer duration. These problems are overcome in the stable transformation approach in plants. For example, in *Arabidopsis* the most widely used in planta transformation (Clough and Bent 1998) method is a convenient and very efficient process for generating stable transgenic plants.

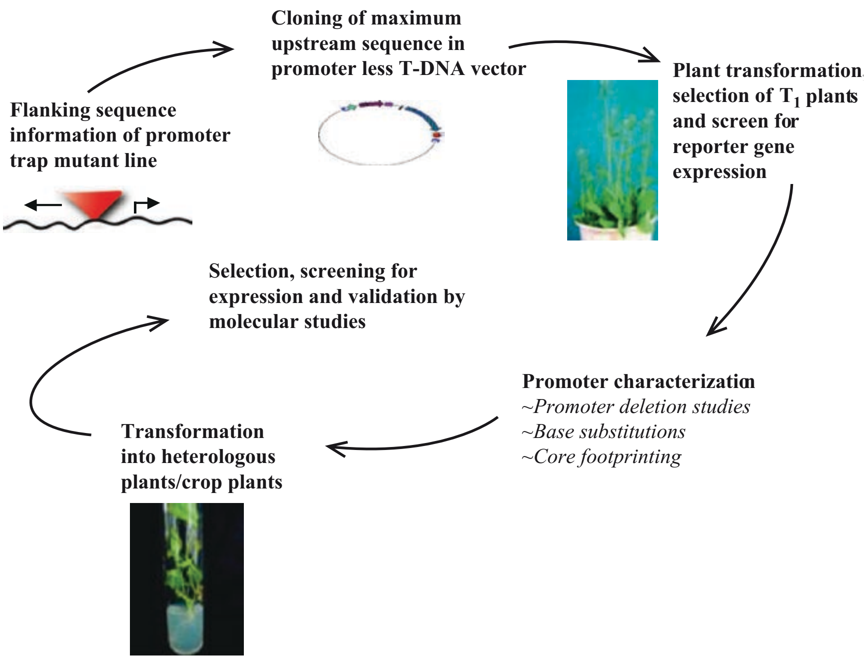


Fig. 23.6 An overview of transgenic-based characterization of putative plant promoter. The steps involve cloning of the identified promoter element from the native plant genome, analysis and validation of the regulatory elements conferring the expression specificity

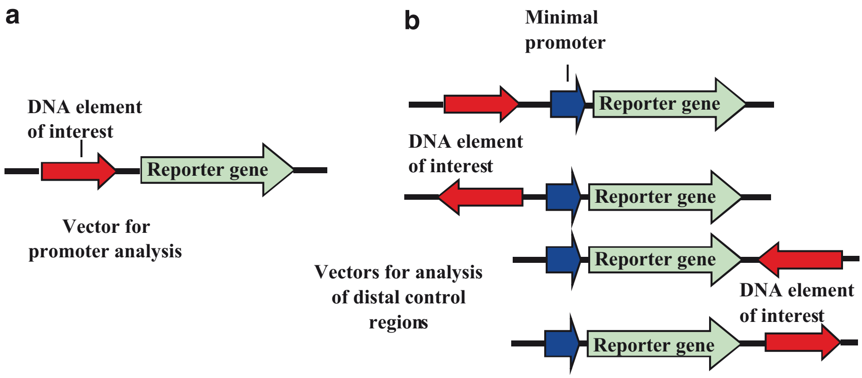


Fig. 23.7 Vectors for identification of regulatory elements in a DNA fragment. (**a**) Construct design for analyzing a DNA fragment for the identification of regulatory elements. (**b**) Different vector strategies for identification and validation of the distal regulatory elements like enhancers in a promoter. Only the placement of the DNA region to be analyzed with respect to the reporter gene is depicted in the constructs

23.6.3.1 Deletion-Based Promoter Analysis

Promoter deletion analysis is one of the primary and most widely employed techniques in promoter analysis. In principle, the technique involves a series of truncated promoter fragments fused to reporter gene and their activity pattern in transgenic plants is detected to identify functional regulatory modules. The series of promoter deletion fragments are generated either through restriction enzyme digestions from the already cloned full-length promoter or through serial digestion of the putative promoter sequence from the 5′ end using Exonuclease III (Fig. 23.8a). The uniformity in the size of the deletion fragments is maintained by timely removal of aliquots from the reactions. The protruding ends of the promoter fragments are polished with S1 nuclease before cloning. One of the most easy and popular strategies to generate promoter deletion constructs is to employ PCR with predefined primer sets to amplify subset of promoter fragments (Fig. 23.8b). The amplified products are cloned into plant transformation vector to generate a promoter-reporter fusion construct. In recent years, PCR based deletion approach is being increasingly employed for the characterization of plant promoters (Kato et al. 2005; Martinez-Trujillo et al. 2004; Ng et al. 2004; Schunmann et al. 2004; Zheng et al. 2004).

23.6.3.2 Linker Scanning for Regulatory Element Identification

The linker scanning approach involves creation of a series of mutated sites in the promoter DNA in small clusters. Unlike promoter deletions, the principle behind this technique is to introduce clusters of mutation (usually 6–8 bp) in a promoter

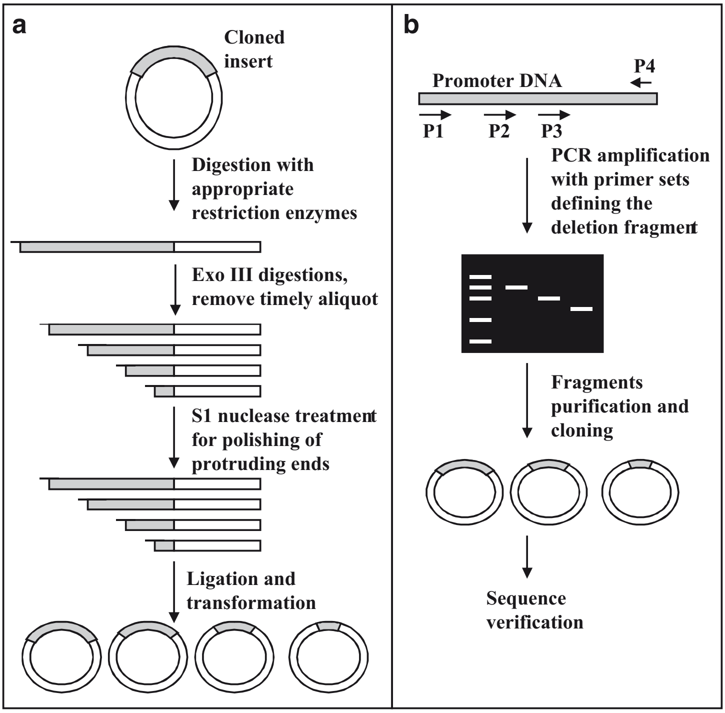


Fig. 23.8 Strategies for promoter deletions and cloning. (**a**) Exonuclease III/S1 nuclease mediated deletion; (**b**) PCR based promoter deletion. P1-P4 represent primers that define the length of deletion fragments

sequence without reducing the promoter length, in a manner that the position of the promoter elements and the TSS remains unaffected. The mutated site in a construct overlaps with that in another construct. All the individual constructs with mutated sequences are evaluated for their functional efficacy in transgenic plants. In the process, the mutated cluster present in all the constructs altogether scans the entire promoter region for identification of functional regulatory elements (McKnight and Kingsbury 1982).

The PCR-based linker scanning involves a three-step amplification to generate the mutated constructs (Fig. 23.9). The first PCR reaction is performed using a set of 5′ end primers containing point mutations in 6–8 bases in the middle of the primer and with a fixed 3′ end primer. The resultant ladder of amplified products contains the mutated sites. The next step of PCR reaction involves only the 3′ end primer to generate single stranded DNA using the amplified products of the first reaction as templates. The single stranded products serve as the 3′ end primer in the third PCR reaction and pairs with a 5′ end primer to produce full-length promoter fragments with mutated sequences (Li and Shapiro 1993). For example, linker scanning along with promoter deletion have been successfully employed in plants to identify *cis*-acting regulatory elements in a number of cases for identification of

regulatory elements of *starch-binding enzyme1* (*SBE*), promoter of maize endosperm (Kim and Guiltinan 1999) and *PR-1* promoter of *Arabidopsis thaliana* (Lebel et al. 1998)

23.6.3.3 Base Substitution Mutation for Validating Regulatory Elements

Mutation by base substitution in a promoter DNA is an important functional assay of a promoter to identify and validate the key regulatory elements present in the region. A two-step PCR reaction is employed to generate mutations in different

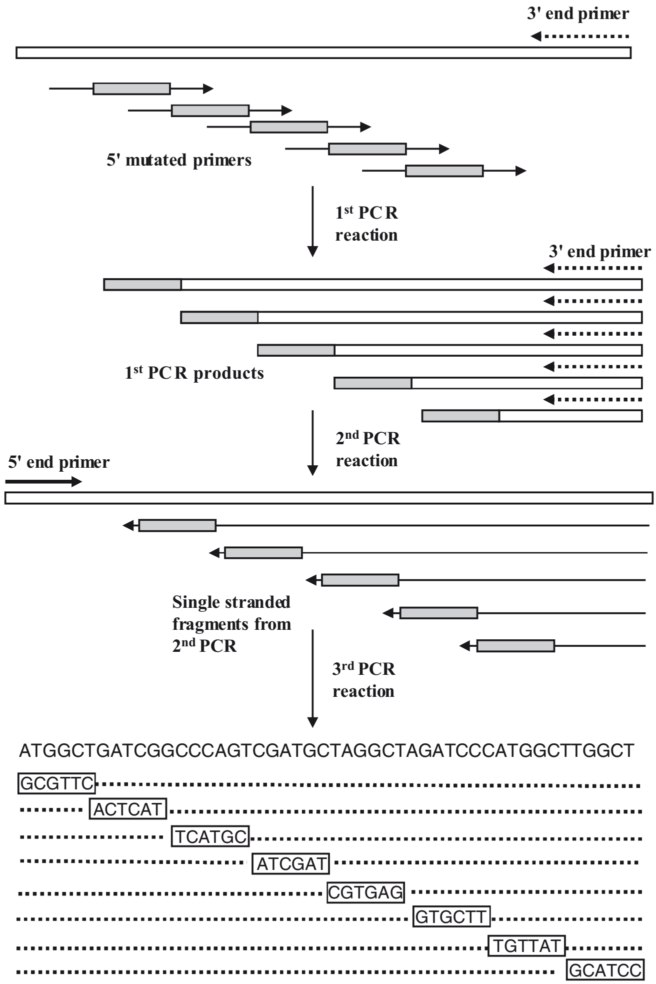


Fig. 23.9 A schematic diagram of PCR based linker scanning strategy. Three PCR reactions are performed to generate constructs with mutated sequence (adapted from Li and Shapiro 1993)

constructs (Fig. 23.10). The strategy involves designing of two internal primers spanning the region to be mutated, such that they complement in their 5′ end and consists of the mutated site. In the first PCR reaction the internal primers are paired separately with two other flanking primers, which specify the two ends of the promoter DNA to be analyzed. The two amplified fragments overlap at one end in their mutated sequence. The two PCR products are denatured and allowed to hybridize at the overlapping region followed by end filling with *Taq* polymerase. The resulting fragment generates a base substituted mutant construct. The base substitution can be designed only to target selected regions of a promoter presumed to contain important regulatory sequences or can be employed to span a whole promoter DNA. The resulting mutated constructs are sub cloned into plant transformation vector for transgenic analysis of promoter activity. Few examples of employing base substitution mutagenesis in characterization of plant promoters for interacting *cis*-acting elements involve identification of *ABRE* and *DRE* elements in the *rd29A* gene promoter of *Arabidopsis* (Narusaka et al. 2003) and regulatory elements responding to ABA in barley (Shen et al. 1996).

23.6.3.4 Detection of Enhancer Activity

The defining features of enhancer elements, namely, their ability to function in orientation and location independent manner are the basis of designing constructs to identify and validate presence of enhancers present in a DNA fragment (Blackwood and Kadonaga 1998). Enhancer element can activate a minimal promoter sequence (truncated *35S* CaMV) (Vorst et al. 1993). For identification of enhancer element, the fragment in question is cloned separately in both the orientations (i.e., forward and reverse) with respect to the reporter gene with minimal promoter (Fig. 23.7) (Jorgensen et al. 1991). The positional effect of the enhancer element is detected by cloning it from a large distance and even in downstream, with respect to the reporter gene (Anthonisen et al. 2002; Vorst et al. 1993). Since enhancer elements are known to control the level of gene transcription, quantification of the reporter protein activity in different constructs is an important parameter to identify a transcriptional enhancer element within a promoter (Anthonisen et al. 2002; Guo et al. 2004; Yamagata et al. 2002). The tissue-specificity of an enhancer element is identified by expression pattern of reporter gene in specific cell types and tissues in transgenic plants (Campisi et al. 1999; Yang et al. 2005).

23.6.3.5 Analysis of TF Binding Sites in Plant Promoter

Transcription factors (TFs) are DNA binding proteins, which bind to specific regulatory elements of a promoter in a sequence-specific manner. The TFs interact to these *cis*-sequences in a combinatorial manner in imparting spatial and temporal regulation of a gene (Singh 1998). Therefore, identification of TFs interacting with a promoter region is crucial in deciphering the transcriptional regulation.

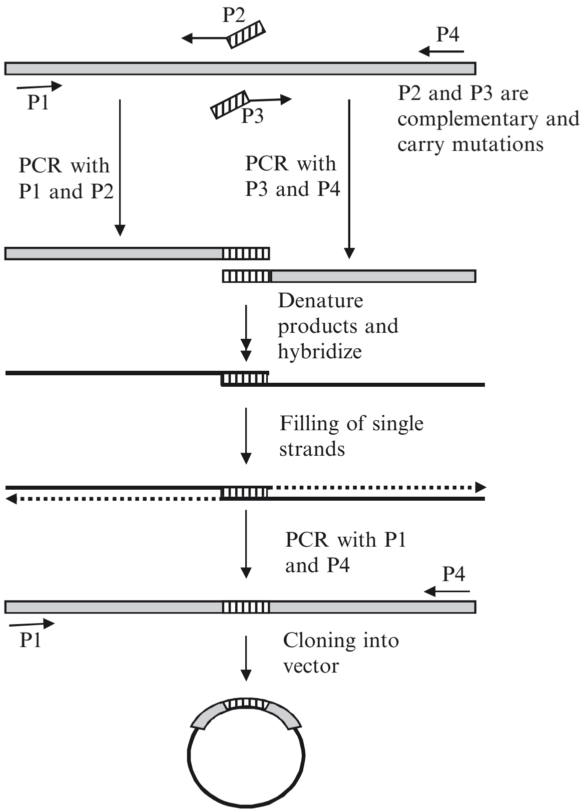


Fig. 23.10 Vector construction strategy of base substituted promoter fragments. P1–P4 represents the primers (adapted from Carey and Smale 2000)

Identification of candidate TFs are possible through experimental approaches like electrophoretic mobility shift assay (EMSA) and foot printing techniques using DNaseI and dimethyl sulphate (DMS).

The electrophoretic mobility shift assay (EMSA) or gel shift assay is the one of the methods for detecting whether a given DNA sequence interactes with proteins. The technique involves a labeled promoter sequence incubated with cognate DNA-binding proteins or total nuclear extract, which are then subjected to electrophoresis in a non-denaturing polyacrylamide gel alongside an untreated DNA sequence. The evidence of protein binding is indicated by the differences in the electrophoretic mobility of the protein bound and free DNA molecules. The protein-bound DNA migrates slower than the free DNA. The EMSA as a tool has been employed in several plant promoter analyses (Abe et al. 2001; Hettiarachchi et al. 2003; Kim and Guiltinan 1999; Meister et al. 2004; Tucker et al. 2002). A more precise approach of identifying TFs binding to *cis*-regulatory sequences is carried out by footprinting techniques using DNaseI. In DNaseI footprinting, similar to EMSA, the promoter

DNA is labeled and allowed to interact with DNA-binding factors. The protein bound DNA is subjected to DNase1 digestion followed by electrophoresis on a denaturing grade sequencing gel. The resultant ladder of fragments is compared with that of untreated promoter DNA cleaved with DNase1. Since, protein bound to promoter DNA prevents nucleolytic cleavage at the site where protein is bound; the comparison of the banding pattern of the cleavage products obtained on sequencing gels reveals the regional sequence of DNA involved in binding to proteins. In the sequencing gel bands corresponding to the region wherein protein is bound will not be present as they would not be subjected to nucleolytic cleavage. The DNaseI mediated footprinting method is mostly performed under *in vitro* condition and has been employed to identify regulatory elements of several plant promoters (Kosugi et al. 1995; Ohgishi et al. 2001). The DNase I footprinting methods described is performed '*in vitro*'. However, the sites identified may not necessarily correspond with the actual TF binding under *in vivo* conditions. The reason is that the TFs may be absent from a particular tissue where the gene is active or it may be activated under a different signaling pathway.

The DMS mediated footprinting technique, instead of DNaseI, utilizes dimethyl sulphate to methylate the 'G' nucleotides of a DNA sequence, followed by cleavage at the methylated bases with piperdine. The cleaved products are compared with that of untreated promoter DNA on denaturing sequencing gel. The TFs bound to regulatory sequences of a promoter prevents methylation and thus a comparison of the banding pattern of cleaved products of protein bound DNA with the free DNA reveals the region to which proteins are bound. In contrast to DNaseI footprinting, which is used in vitro, DMS based method can be used both *in vivo* and *in vitro*. The advantage of using DMS over DNase1 in plant is that unlike DNaseI, DMS can readily diffuse into the intact plant tissues and does not require tedious and risky procedure of nuclei isolation from the cells (Busk and Pages 2002). The additional advantage of the *in vivo* footprinting technique, also known as 'genomic footprinting' (Busk and Pages 2002), is that the tool can investigate even a complex binding of multiple TFs to a single promoter DNA as demonstrated by analyzing the β-phaseolin promoter during embryogenesis (Li and Hall 1999). The DMS mediated genomic footprinting was employed to detect binding of nuclear factors in regulatory elements of several plant promoters (Lebel et al. 1998; Opsahl-Sorteberg et al. 2004; Paul and Ferl 1991).

The ligation mediated PCR (LM-PCR) is another useful tool used in both DNaseI and DMS mediated genomic footprinting(Carey and Smale 2000; Dai et al. 2000) (Fig. 23.11). After treatment with either DNaseI or DMS and the cleavage, the technique involves, three rounds of PCR amplification of the target DNA using three gene specific primers at one-end and oligonucleotide linker and linker specific primer at the other end. The first PCR reaction involves a single gene specific primer that generates blunt ended PCR products up to the cleavage site of the DNA. The linker molecules are attached to the PCR products followed by a second PCR reaction using a linker specific primer paired with another gene specific primer slightly internal to the first. The amplified product is used as template in the final PCR reaction in which the third gene specific primer is end labeled and paired with

linker specific primer. The amplified products are analyzed on the sequencing gel to resolve the footprint bands (Dimitrova et al. 1994).

23.6.3.6 Efficacy of Promoter Elements in Heterologus Plant Systems

Although, *A. thaliana* is extensively used in analyzing plant promoters through transgenic studies (Pereira 2000) the practical application of the promoters requires their validation in the heterologus target plant system. The functional analysis of an *Arabidopsis* plastocyanin promoter in transgenic tobacco plant has resulted into identification of a positive regulatory element conferring light and chloroplast dependent gene expression. The findings correlate with the organ specific expression of the plastocyanin gene in *Arabidopsis* (Vorst et al. 1993). Similarly the barley aleurone specific *ltp2* (lipid transfer protein) gene promoter was found to produce well-conserved expression pattern in the aleurone tissues of transgenic rice (Kalla 1994; Opsahl-Sorteberg 2004).

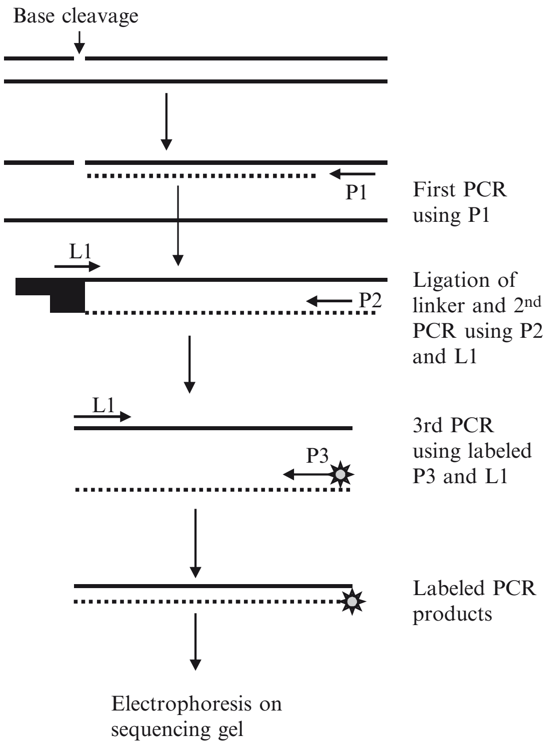


Fig. 23.11 A schematic diagram of the LM-PCR strategy used for in vivo foot printing. P1–P3 represents gene specific primers; L1 represents linker specific primer. The labeling of the DNA fragment is represented with a star (adapted from Carey and Smale 2000)

Plant promoters exhibiting transcriptional activity in *A. thaliana* may not necessarily remain active in other plant systems partly because of the absence of the TFs or different signaling cues. For example, auxin-inducible *ACC* gene promoter in a *Vigna radiata* (*VR-ACS1*) was induced only when exposed to stimuli like auxin. The same *VR-ACS1* promoter in transgenic *Arabidopsis* and tobacco exhibited strong constitutive expression throughout the development but failed to respond to auxin in heterologous host (Cazzonelli et al. 2005). Similarly the wound inducible peroxidase gene promoters *poxA* and *poxN*, which drive gene expression in a tissue specific manner in rice were found to be ineffective in the heterologous tobacco system (Ito et al. 2000). Thus any plant promoter has to be evaluated for its efficacy on a case to case basis, in a target plant.

23.7 Conclusions

The recent past has witnessed a surge of new empirical and theoretical approaches in the field of gene and promoter isolations from plants. The T-DNA based promoter entrapment strategy represents a specialized tool to generate large collections of random insertion mutants with specific cell-type markers and facilitating easy screening of reporter gene expression in variety of tissues/cell types and in response to a myriad of environmental cues. T-DNA promoter trapping can effectively detect the promoters of the functionally redundant genes and those whose activities are restricted to spatial and temporal regulations only during the plant developmental stages, including the presence of silent and cryptic elements which are otherwise not associated with any expressed gene. The increasing availability of genomic sequence information in several plant species and the advancements in the area of functional genomics techniques to identify and characterize the tagged regulatory sequences has made T-DNA tagging a powerful tool to be employed in several plant species. The validation of the plant promoter in heterologous target plant system remains an important prerequisite towards deployment of novel plant promoter elements in transgenics both for applied as well as basic research.

In spite of the fact that considerable developments have taken place in T-DNA based promoter trapping techniques, there is still a lot of scope for further improvement. Until now, majority of the promoter trapping vectors are based on GUS reporter gene that requires a destructive method for detecting the enzyme activity. Further the detection of the enzyme activity requires the addition of a substrate of the GUS. These are particularly serious limitations for identifying regulatory elements conferring expression in deep-seated cells/ tissues or where the expression is restricted to a limited number of cells. The detection of the enzyme depends upon the infiltration of substrate, time of incubation and other assay condition. The option of nondestructive reporter gene like green fluorescent protein (gfp) is emerging as an alternative strategy for identifying promoter activity. Since the fluorescence of gfp is intrinsic and does not require any additional factors. The application of promoter trapping tool in the plants beyond *Arabidopsis* would certainly lead to the isolation of regulatory elements from diverse plant sources.

Acknowledgements We gratefully acknowledge the financial assistance provided under the NATP, CGP Grant (CGPII/253) for promoter trapping research in *Arabidopsis* in our laboratory. The study leave grant of IINRG (ICAR) to DS is gratefully acknowledged.

References

Abe M, Takahashi T, Komeda Y (2001) Identification of a *cis*-regulatory element for L1 layer specific gene expression, which is targeted by an L1-specific homeodomain protein. Plant J 26:487–494

Albani D, Sardana R, Robert LS, Altosaar I, Arnison PG, Fabijanski SF (1992) A *Brassica napus* gene family, which shows sequence similarity to ascorbate oxidase, is expressed in developing pollen. Molecular characterization and analysis of promoter activity in transgenic tobacco plants. Plant J 2:331–342

Alvarado MC, Zsigmond LM, Kovacs I, Cseplo A, Koncz C, Szabados L (2004) Gene trapping with firefly luciferase in *Arabidopsis*. Tagging of stress-responsive genes. Plant Physiol 134:18–27

Andre D, Colau D, Schell J, van Montagu M, Hernalsteens J-P (1986) Gene tagging in plants by a T-DNA insertion that generates *aph (3′)II* plant gene fusions. Mol Gen Genet 204: 512–518

Anthonisen IL, Kasai S, Kato K, Salvador ML, Klein U (2002) Structural and functional characterization of a transcription-enhancing sequence element in the *rbcL* gene of the *Chlamydomonas* chloroplast genome. Curr Genet 41:349–356

Azpiroz-Leehan R, Feldmann KA (1997) T-DNA insertion mutagenesis in *Arabidopsis*: going back and forth. Trends Genet 13:152–156

Bade J, van Grinsven E, Custers J, Hoekstra S, Ponstein A (2003) T-DNA tagging in *Brassica napus* as an efficient tool for the isolation of new promoters for selectable marker genes. Plant Mol Biol 52:53–68

Barthels N, van der Lee FM, Klap J, Goddijn OJ, Karimi M, Puzio P, Grundler FMW, Ohl SA, Lindsey K, Robertson WM, Montagu MV, Gheysen G, Sijmons PC (1997) Regulatory sequences of *Arabidopsis* drive reporter gene expression in nematode feeding structures. Plant Cell 9:2119–2134

Beals TP, Goldberg RB (1997) A novel cell ablation strategy blocks tobacco anther dehiscence. Plant Cell 9:1527–1545

Bevan MW, Flavel RB, Clitton MD (1983) A chimeric antibiotic resistance gene as a selectable marker for plant cell transformation. Nature 304:184–187

Bevan M, Walsh S (2005) The Arabidopsis genome: a foundation for plant research. Genome Res 15:1632–1642

Blackwood EM, Kadonaga JT (1998) Going the distance: a current view of enhancer action. Science 281:61–63

Blanchette M, Sinha S (2001) Separating real motifs from their artifacts. Bioinformatics 17:30–38

Blanchette M and Tompa T (2003) FootPrinter: a program designed for phylogenetic footprinting. Nucleic Acids Research, 31: 3840–3842

Busk PK, Pages M (2002) *In vivo* footprinting of plant tissues. Plant Mol Biol Rep 20:287–297

Buzas DM, Lohar D, Sato S, Nakamura Y, Tabata S, Vickers CE, Stiller J, Gresshoff PM (2005) Promoter trapping in *Lotus japonicus* reveals novel root and nodule GUS expression domains. Plant Cell Physiol 46:1202–1212

Campisi L, Yang Y, Yi Y, Heilig E, Herman B, Cassista AJ, Allen DW, Xiang H, Jack T (1999) Generation of enhancer trap lines in *Arabidopsis* and characterization of expression patterns in the inflorescence. Plant J 17:699–707

Carey MF, Smale ST (2000) Transcription initiation site mapping. In: Transcriptional regulation in eukaryotes: concepts, strategies and techniques. CSHL, New York, pp 97–136

Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, Bayerlein M, Werner T (2005) MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics 21:2933–2942

Casson SA, Chilley PM, Topping JF, Evans IM, Souter MA, Lindsey K (2002) The *POLARIS* gene of *Arabidopsis* encodes a predicted peptide required for correct root growth and leaf vascular patterning. Plant Cell 14:1705–1721

Cazzonelli CI, McCallum EJ, Lee R, Botella JR (2005) Characterization of a strong, constitutive mung bean (*Vigna radiata* L.) promoter with a complex mode of regulation *in planta*. Transgenic Res 14:941–967

Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J 16:735–743

Custers JHHV, Melchers LS, Tigelaar H, Bade JB, Spiegeler JJM, van der Meijs PJ, Simons BH, Stuiver MH (2002) T-DNA tagging of a pathogen inducible promoter in *Arabidopsis thaliana*. Mol Plant Pathol 3:39–249

Dai SM, Chen HH, Chang C, Riggs AD, Flanagan SD (2000) Ligation-mediated PCR for quantitative in vivo footprinting. Nat Biotechnol 18:1108–1111

Datla R, Anderson JW, Selvaraj G (1997) Plant promoters for transgene expression. Biotechnol Ann Rev 3:269–296

Davuluri RV, Sun H, Palaniswamy SK, Matthews N, Molina C, Kurtz M, Grotewold E (2003) AGRIS: *Arabidopsis* gene regulatory information server, an information resource of *Arabidopsis cis*-regulatory elements and transcription factors. BMC Bioinformatics 23:25

De Greve H, Nguyen VK, Deboeck F, Thia-Toong L, Karimi M, Hernalsteens JP (2001) T-DNA tagging of the translation initiation factor elF-4A1 of *Arabidopsis thaliana*. Plant Sci 161:685–693

Digeon JF, Guiderdoni E, Alary R, Michaux-Ferriere N, Joudrier P, Gautier MF (1999) Cloning of a wheat *puroindoline* gene promoter by IPCR and analysis of promoter regions required for tissue-specific expression in transgenic rice seeds. Plant Mol Biol 39:1101–1112

Dilkes BP, Feldmann KA (1998) Cloning genes from T-DNA tagged mutants. In: Martinez Zapater JM, Salinas J (eds) Arabidopsis Protocols. Humana, Totowa, NJ, pp 339–351

Dimitrova D, Giacca M, Falaschi A (1994) A modified protocol for *in vivo* footprinting by ligation-mediated polymerase chain reaction. Nucleic Acids Res 22:532–533

Engineer CB, Fitzsimmons KC, Schmuke JJ, Dotson SB, Kranz RG (2005) Development and evaluation of a Gal4-mediated LUC/GFP/GUS enhancer trap system in *Arabidopsis*. BMC Plant Biology. 5:9. doi:10.1186/1471-2229-5-9

Filichkin SA, Wu Q, Busov V, Meilan R, Lanz-Garcia C, Groover A, Goldfarb B, Ma C, Dharmawardhana P, Brunner A, Strauss SH (2006) Enhancer trapping in woody plants: Isolation of the ET304 gene encoding a putative AT-hook motif transcription factor and characterization of the expression patterns conferred by its promoter in transgenic *Populus* and *Arabidopsis*. Plant Sci 171:206–216

Galuschka C, Schindler M, Bülow L, Hehl R (2007) AthaMap web tools for the analysis and identification of co-regulated genes. Nucleic Acids Res 35:D857–D862

Goddijn OJM, Lindsey K, Lee FM, van der Klap JC, Sijmons PC (1993) Differential expression in nematode-induced feeding structures of transgenic plants harboring promoter-GUSA fusion constructs. Plant J 4:863–873

Goldsbrough A, Bevan M (1991) New patterns of gene activity in plants detected using *Agrobacterium* vector. Plant Mol Biol 16:263–269

Groover A, Fontana JR, Dupper G, Ma C, Martienssen R, Strauss S, Meilan R (2004) Gene and enhancer trap tagging of vascular-expressed genes in poplar trees. Plant Physiol 134:1742–1751

Guo H, Chen X, Zhang H, Fang R, Yuan Z, Zhang Z, Tian Y (2004) Characterization and activity enhancement of the phloem-specific pumpkin *PP2* gene promoter. Transgenic Res 13:559–566

Guo A, He K, Liu D, Bai S, Gu X, Wei L, Luo J (2005) DATF: a Database of *Arabidopsis* Transcription Factors. Bioinformatics 21:2568–2569

Farrar K, Evans IM, Topping JF, Souter MA, Nielsen JE, Lindsey K (2003) *EXORDIUM*-a gene expressed in proliferating cells and with a role in meristem function, identified by promoter trapping in *Arabidopsis*. Plant J 33:61–73

Favery B, Lecomte P, Gil N, Bechtold N, Bouchez D, Dalmasso A, Abad P (1998) RPE, a plant gene involved in early developmental steps of nematode feeding cells. EMBO J 17:6799–6811

Fedoroff NV, Smith DL (1993) A versatile system for detecting transposition in *Arabidopsis*. Plant J 3:273–289

Fickett JW, Hatzigeorgiou AG (1997) Eukaryotic promoter recognition. Genome Res 7:861–878

Fobert PR, Miki BL, Iyer VN (1991) Detection of gene regulatory signals in plants revealed by T-DNA-mediated fusions. Plant Mol Biol 17:837–851

Fobert PR, Labbe H, Cosmopoulos J, Gottllob-McHugh S, Ouellet T, Hattori J, Sunohara G, Iyer VN, Miki B (1994) T-DNA tagging of a seed coat-specific cryptic promoter in tobacco. Plant J 6:567–577

Foster E, Hattori J, Labbe H, Ouellet J, Fobert PR, James, LE, Iyer VN, Miki BL (1999) A tobacco cryptic constitutive promoter *tCUP* revealed by T-DNA tagging. Plant Mol Biol 41:45–55

Fraley RT, Rogers SG, Horsch R, Sanders PR, Flick JS, Adams SP, Bittner ML, Brand LA, Fink C, Fry J, Galluppi G, Goldberg S, Hoffmann NL, Woo S (1983) Expression of bacterial genes in plants. Proc Natl Acadof Sci U S A 80:4803–4807

Francis KE, Spiker S (2005) Identification of *Arabidopsis thaliana* transformants without selection reveals a high occurrence of silenced T-DNA integrations. Plant J 41:464–477

Goujon T, Sibout R, Pollet B, Maba B, Nussaume L, Bechtold N, Lu F, Ralph J, Mila I, Barrière Y, Lapierre C, Jouanin L (2003) A new *Arabidopsis thaliana* mutant deficient in the expression of *O*-methyltransferase impacts lignins and sinapoyl esters. Plant Mol Biol 51:973–989

Handberg K, Stougaard J (1992) *Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics. Plant J 2:487–496

Haseloff J, Siemering KR, Prasher DC, Hodge S (1997) Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic *Arabidopsis* plants brightly. Proc Natl Acad Sci U S A 94:2122–2127

Herman L, Jacobs A, van Montagu M, Depicker A (1990) Plant/chromosome marker gene fusion assay for study of normal and truncated T-DNA integration events. Mol Gen Genet 224:248–256

Herrera-Estrella L, Depicker A, Montagu MV, Schell J (1983) Expression of chimeric genes transferred into plant cells using a Ti-plasmid-derived vectors. Nature 303:209–213

Hettiarachchi GHCM, Yadav V, Reddy MK, Chattopadhyay S, Sopory SK (2003) Light-mediated regulation defines a minimal promoter region of TOP2. Nucleic Acids Res 31:5256–5265

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant *cis*-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res 27:297–300

Ito H, Hiraga S, Tsugawa H, Matsui H, Honma M, Otsuki Y, Murakami T, Ohashi Y (2000) Xylem-specific expression of wound-inducible rice peroxidase genes in transgenic plants. Plant Sci 155:85–100

Janssen BJ, Gardner RC (1990) Localized transient expression of GUS in leaf discs following cocultivation with *Agrobacterium*. Plant Mol Biol 14:61–72

Johansson AM, Wang C, Stenberg A, Hertzberg M, Little C, Olsson A (2003) Characterization of a *PttRPS18* promoter active in the vascular cambium region of hybrid aspen. Plant Mol Biol 52:317–329

Jorgensen J-E, Stougaard J, Marcke KA (1991) A two-component nodule-specific enhancer in the soybean *N23* gene promoter. Plant Cell 3:819–827

Kalla R, Shimamoto K, Potter R, Nielsen PS, Linnestad C, Olsen OA (1994) The promoter of the barley aleurone-specific gene encoding a putative 7 kDa lipid transfer protein confers aleurone cell-specific expression in transgenic rice. Plant J 6:849–860

Kapila J, De Rycke R, van Montagu M, Angenon G (1997) An *Agrobacterium*-mediated transient gene expression system for intact leaves. Plant Sci 122:101–108

Kato A, Endo M, Kato H, Saito T (2005) The antisense promoter of *AtRE1*, a retrotransposon in *Arabidopsis thaliana*, is activated in pollens and calluses. Plant Sci 168:981–986

Kertbundit S, De Greve H, Deboeck F, van Montagu M, Hernalsteens J-P (1991) *In-vivo* random β-glucuronidase gene fusions in *Arabidopsis thaliana*. Proc of Natl Acad Sci U S A 88:5212–5216

Kim K-N, Guiltinan MJ (1999) Identification of *cis*-acting elements important for expression of the starch-branching enzyme I gene in maize endosperm. Plant Physiol 121:225–236

Klimyuk VI, Nussaume L, Harrison K, Jones JDG (1995) Novel GUS expression patterns following transposition of an enhancer trap *Ds* element in *Arabidopsis*. Mol Gen Genet 249:357–365

Ko S, Kamada H (2002) Enhancer-trapping system for somatic embryogenesis in carrot. Plant Mol Biol Rep 20:421a–421j

Koncz C, Martini M, Mayerhofer R, Koncz-Kalman Z, Korber H, Redei G, Schell J (1989) High frequency T-DNA tagging in plants. Proc Natl Acad Sci U S A 86:8467–8471

Koltunow AM, Truettner J, Cox KH, Wallroth M, Goldberg RB (1990) Different temporal and spatial gene expression patterns occur during anther development. Plant Cell 2: 1201–1224

Kosugi S, Suzuka I, Ohashi Y (1995) Two of three promoter elements identified in a rice gene for proliferating cell nuclear antigen are essential for meristematic tissue-specific expression. Plant J 7:877–886

Krysan PJ, Young JC, Sussman MR (1999) T-DNA as insertional mutagen in *Arabidopsis*. Plant Cell 11:2283–2290

Kuhlemeier C, Green PJ, Chua NM (1987) Regulation of gene expression in higher plants. Ann Rev Plant Physiol 38:221–257

Laplaze L, Parizot B, Baker A, Ricaud L, Martiniere A, Auguy F, Franche C, Nussaume L, Bogusz D, Haseloff J (2005) GAL4-GFP enhancer trap lines for genetic manipulation of lateral root development in *Arabidopsis thaliana*. J Exp Bot 56:2433–2442

Lebel E, Heifetz P, Thorne L, Uknes S, Ryals J, Ward E (1998) Functional analysis of regulatory sequences controlling PR-1 gene expression in *Arabidopsis*. Plant J 16:223–233

Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S (2002) PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. Nucleic Acids Res 30:325–327

Li G, Hall TC (1999) Footprinting *in vivo* reveals changing profiles of multiple factor interactions with the *β-phaseolin* promoter during embryogenesis. Plant J 18:633–641

Li GF, Hall TC, Holmes-Davis R (2002) Plant chromatin: development and gene control. Bioessays 24:234–243

Li XM, Shapiro LJ (1993) Three-step PCR mutagenesis for 'linker scanning'. Nucleic Acids Res 21:3745–3748

Lindsey K, Wei W, Clarke MC, McArdle HF, Rooke LM, Topping JF (1993) Tagging genomic sequences that direct transfer expression by activation of a promoter trap in plants. Trans Res 2:33–47

Liu YG, Mitsukawa N, Oosumi T, Whittier RF (1995) Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. Plant J 8:457–463

Mandal A, Sandgen M, Holmstrom KO, Gallois P, Palua ET (1995) Identification of *Arabidopsis thaliana* sequences responsive to low temperature and abscisic acid by T-DNA tagging and *in vivo* gene fusion. Plant Mol Biol Rep 13:243–254

Marks MD, Feldmann KA (1989) Trichome development in *Arabidopsis thaliana*. *L.* T-DNA tagging of the *GLABROUS1* gene. Plant Cell 1:1043–1050

Martınez-Trujillo M, Chávez-Bárcenas T, Limones-Briones V, Simpson J, Herrera-Estrella L (2004) Functional analysis of the promoter of the rice *sucrose phosphate synthase* gene (*sps1*). Plant Sci 166:131–140

Martirani L, Stiller J, Mirabella R, Alfano F, Lamberti A, Radutoiu SE, Iaccarino M, Gresshoff PM, Chiurazzi M (1999) T-DNA tagging of nodulation and root related genes in *Lotus japonicus*: expression patterns and potential for promoter trapping and insertional mutagenesis. Mol Plant Microbe Interact 12:275–284

Matys V, Fricke E, Geffers R, Gößling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos D-U, Land S, Lewicki-Potapov B, Michael H, Münch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E (2003) TRANSFAC®: transcriptional regulation, from patterns to profiles. Nucleic Acids Res 31:374–378

McKnight SL, Kingsbury R (1982) Transcriptional control signals of a eukaryotic protein-coding gene. Science 217:316–324

Meissner R, Chague V, Zhu Q, Emmanuel E, Elkind Y, Levy AA (2000) A high throughput system for transposon tagging and promoter trapping in tomato. Plant J 22:265–274

Meister RJ, Williams LA, Monfared MM, Gallagher TL, Kraft EA, Nelson CG, Gasser CS (2004) Definition and interaction of a positive regulatory element of the *Arabidopsis INNER NO OUTER* promoter. Plant J 37:426–438

Michael TP, McClung CR (2003) Enhancer trapping reveals widespread circadian clock transcriptional control in *Arabidopsis*. Plant Physiol 132:629–639

Mishra RN, Singla-Pareek SL, Nair S, Sopory SK, Reddy MK (2002) Directional genome walking using PCR. Biotechniques 33:830–832

Molina C, Grotewold E (2005) Genome wide analysis of *Arabidopsis* core promoters. BMC Genomics 6:25 (online) http://www.biomedcentral.com/1471-2164/6/25

Mollier P, Montoro P, Delarue M, Bechtold N, Bellini C, Pelletier G (1995) Promoterless *gusA* expression in a large number of *Arabidopsis thaliana* transformants obtained by *in planta* infiltration method. CR Acad Sci Paris 318:465–474

Mollier P, Hoffman B, Orsel M, Pelletier G (2000) Tagging of a cryptic promoter that confers root-specific GUS expression in *Arabidopsis thaliana*. Plant Cell Rep 19:1076–1083

Mudge SR, Birch RG (1998) T-DNA tagging and characterization of a novel meristem-specific promoter from tobacco. Aust J Plant Physiol 25:637–643

Nacry P, Camilleri C, Courtial B, Caboche M, Bouchez D (1998) Major chromosomal rearrangements induced by T-DNA transformation in *Arabidopsis*. Genetics 149:641–650

Narusaka Y, Nakashima K, Shinwari ZK, Sakuma Y, Furihata T, Abe H, Narusaka M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Interaction between two *cis*-acting elements ABRE and DRE, in ABA-dependent expression of *Arabidopsis rd29A* gene in response to dehydration and high-salinity stresses. Plant Journal 34:137–148

Ng DWK, Chandrasekharan MB, Hall TC (2004) The 5' UTR negatively regulates quantitative and spatial expression from the *ABI3* promoter. Plant Mol Biol 54:25–38

Ochman H, Gerber AS, Hartl DL (1988) Genetic applications of an inverse polymerase chain reaction. Genetics 120:621–623

O'Connor TR, Dyreson C, Wyrick JJ (2005) Athena: a resource for rapid visualization and systematic analysis of Arabidopsis promoter sequences. Bioinformatics 21:4411–4413

Ohgishi M, Oka A, Morelli G, Ruberti I, Aoyama T (2001) Negative autoregulation of the *Arabidopsis* homeobox gene *ATHB-2*. Plant J 25:389–398

Okresz L, Mathe C, Horvath E, Schell J, Koncz C, Szabados L (1998) TDNA trapping of a cryptic promoter identifies an ortholog of highly conserved SNZ growth arrest response genes in *Arabidopsis*. Plant Sci 138:217–228

Opsahl-Sorteberg H-G, Divon HH, Nielsen PS, Kalla R, Hammond-Kosack M, Shimamoto K, Kohli A (2004) Identification of a 49-bp fragment of the *HvLTP2* promoter directing aleurone cell specific expression. Gene 341:49–58

Osborne BI, Baker B (1995) Movers and shakers: maize transposons as tools for analyzing other plant genomes. Curr Opin Cell Biol 7:406–413

Palaniswamy SK, James S, Sun H, Lamb RS, Davuluri, RV, Grotewold E (2006) AGRIS and AtRegNet: a platform to link *cis*-regulatory elements and transcription factors into regulatory networks. Plant Physiol 140:818–829

Parinov S, Sundaresan V (2000) Functional genomics in *Arabidopsis*: large-scale insertional mutagenesis complements the genome sequencing project. Curr Opin Biotechnol 11:157–161

Parinov S, Sevugan M, Ye D, Yang W-C, Kumaran M, Sundaresan V (1999) Analysis of flanking sequences from *Dissociation* insertion lines: a database for reverse genetics in *Arabidopsis*. Plant Cell 11:2263–2270

Paul AL, Ferl RJ (1991) *In vivo* footprinting reveals unique *cis*-elements and different modes of hypoxic induction in maize *Adh1* and *Adh2*. Plant Cell 3:159–168

Pereira A (2000) A transgenic perspective on plant functional genomics. Trans Res 9:245–260

Plesch G, Kamann E, Mueller-Roeber B (2000) Cloning regulatory sequences mediating guard-cell-specific gene expression. Gene 249:83–89

Prasad AM, Sivanandan C, Resminath R, Thakare DR, Bhat SR, Srinivasan R (2005) Cloning and characterization of a pentatricopeptide protein-encoding gene (*LOJ*) that is specifically expressed in lateral organ junctions in *Arabidopsis thaliana*. Gene 353:67–79

Puzio PS, Lausen J, Almeida-Engler J, Cai D, Gheysen G, Grundler FMW (1999) Isolation of a gene from *Arabidopsis thaliana* related to nematode feeding structures. Gene 239:163–172

Puzio PS, Lausen J, Heinen P, Grundler FM (2000) Promoter analysis of *pyk20*, a gene from *Arabidopsis thaliana*. Plant Sci 157:245–255

Ream LW, Gordon MP (1982) Crown gall disease and prospects for genetic manipulation of plants. Science 218:854–859

Remy S, Thiry E, Coemans B, Windelinckx S, Swennen R, Sagi L (2005) Improved T-DNA vector for tagging plant promoters via high-throughput luciferase screening. BioTechniques 38:763–770

Resminath R, Prasad AM, Thakare DR, Sivanandan C, Bhat SR, Srinivasan (2005a) Promoter trapping in *Arabidopsis* using T-DNA insertional mutagenesis. J Plant Biochem Biotechnol 14:1–8

Resminath R, Prasad AM, Srinivasan R (2005b) T-DNA insertional mutagenesis in *Arabidopsis*: a tool for functional genomics. Electron J Biotechnol (Online) 8. http://www.ejbiotechnology.info/content/vol8/issue1/full /4/ index.html. ISSN 0717-3458

Reuber TL, Ausubel FM (1995) Differential mRNA display. Methods Cell Biol 49:431–440

Richmond TA, Somerville CR (2000) The cellulose synthase superfamily. Plant Physiol 124:495–498

Rombauts S, Dehais P, Van Montagu M, Rouze P (1999) PlantCARE, a plant *cis*-acting regulatory element database. Nucleic Acids Res 27:295–296

Rombauts S, Florquin K, Lescot M, Marchal K, Rouze P, van de Peer Y (2003) Computational approaches to identify promoters and *cis*-regulatory elements in pant genomes. Plant Physiol 132:1162–1176

Saha D, Prasad AM, Sujatha TP, Kumar V, Jain PK, Bhat SR, Srinivasan R (2007) *In silico* analysis of the *Lateral Organ Junction* (*loj*) gene and promoter of *Arabidopsis thaliana*. In Silico Biol 7:7–19.

Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res 32:D91–94

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470

Schug J, Overton GC (1997) TESS: Transcription element search software on the WWW. Technical report CBIL-TR-1997-1001-v0.0, Computational Biology and Informatics Laboratory, School of Medicine, University of Pennsylvania

Schunmann PHD, Richardson AE, Vickers CE, Delhaize E (2004) Promoter analysis of the Barley *Pht1;1* phosphate transporter gene identifies regions controlling root expression and responsiveness to phosphate deprivation. Plant Physiology 136:4205–4214

Shen Q, Zhang P, Ho T-HD (1996) Modular nature of abscisic acid (ABA) response complexes: composite promoter units that are necessary and sufficient for ABA induction of gene expression in barley. Plant Cell 8:1107–1119

Siebert PD, Chenchik A, Kellogg DE, Lukyanov KA, Lukyanov SA (1995) An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Res 23:1087–1088

Singh KB (1998) Transcriptional regulation in plants: the importance of combinatorial control. Plant Physiol 118:1111–1120

Sivanandan C, Sujatha TP, Prasad AM, Resminath R, Thakare DR, Bhat SR, Srinivasan R (2005) T-DNA tagging and characterization of a cryptic root-specific promoter in *Arabidopsis*. Biochimica et Biophysica Acta 1731:202–208

Small ID, Peeters N (2000) The PPR motif - a TPR-related motif prevalent in plant organellar proteins. Trends Biochem Sci 25:46–47

Smith DL, Federoff NV (1995) *LRP1*, a gene expressed in expressed in lateral and adventitious root primordial of *Arabidopsis*. Plant Cell 7:735–745

Smith HA, Swaney SL, Parks TD, Wernsman EA, Dougherty WG (1994) Transgenic plant virus resistance mediated by untranslatable sense RNAs: expression, regulation, and fate of nonessential RNAs. Plant Cell 6:1441–1453

Springer PS (2000) Gene traps: tools for plant development and genomics. Plant Cell 12:1007–1020

Stangeland B, Nestestog R, Grini PE, Skrbo N, Berg A, Salehian Z, Mandal A, Aalen RB (2005) Molecular analysis of Arabidopsis endosperm and embryo promoter trap lines: reporter-gene expression can result from T-DNA insertions in antisense orientation, in introns and in intergenic regions, in addition to sense insertion at the 5' end of genes. J Exp Bot 56:2495–2505

Stangeland B, Salehian Z, Aalen R, Mandal A, Olsen OA (2003) Isolation of GUS marker lines for genes expressed in *Arabidopsis* endosperm, embryo, and maternal tissues. J Exp Bot 54:279–290

Steffens NO, Galuschka C, Schindler M, Bülow L, Hehl R (2004) AthaMap: an online resource for *in silico* transcription factor binding sites in the *Arabidopsis thaliana* genome. Nucleic Acids Res 32:D368–372

Stewart CN (2001) The utility of green fluorescent protein in transgenic plants. Plant Cell Rep 20:376–382

Sundaresan V, Springer P, Volpe T, Haward S, Jones JDG, Dean C, Ma H, Martienssen R (1995) Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. Genes Dev 9:1797–1810

Swaminathan K, Yang Y, Grotz N, Campisi L, Jack T (2000) An enhancer trap line associated with a D-**c**lass cyclin gene in *Arabidopsis*. Plant Physiol 124:1658–1667

Takano M, Egawa H, Ikeda JE, Wakasa K (1997) The structures of integration sites in transgenic rice. Plant J 11:353–361

Thakare DR, Resminath R, Chikhale NJ, Bhat SR, Srinivasan (2006) Molecular and histochemical analysis of a T-DNA tagged mutant of *Arabidopsis* exhibiting anther-specific GUS expression. J Plant Biochem Biotechnol 15:7–13

Thijs G, Marchal K, Lescot M, Rombauts S, De Moor B, Rouze P, Moreau Y (2002) A Gibbs sampling method to detect overrepresented motifs in the upstream regions of co-expressed genes. J Comput Biol 9:447–464

Topping JF, Lindsey K (1997) Promoter trap markers differentiate structural and positional components of polar development in *Arabidopsis*. Plant Cell 9:1713–1725

Topping JF, Wei W, Lindsey K (1991) Functional tagging of regulatory elements in the plant genome. Development 112:1009–1019

Topping JF, Agyeman F, Henricot B, Lindsey K (1994) Identification of molecular markers of embryogenesis in *Arabidopsis thaliana* by promoter trapping. Plant J 5:895–903

Tucker ML, Whitelaw CA, Lyssenko NN, Nath P (2002) Functional analysis of regulatory elements in the gene promoter for an abscission-specific cellulase from bean and isolation, expression, and binding affinity of three TGA-type basic leucine zipper transcription factors. Plant Physiol 130:1487–1496

Tyagi AK (2001) Plant genes and their expression. Curr Sci 80:161–169

Ukai H, Ukai-Tadenuma M, Ogiu T, Tsuji H (2002) A new technique to prevent self-ligation of DNA. J Biotechnol 97:233–242

Upadhyaya NM, Zhou X-R, Zhu Q-H, Ramm K, Wu L, Eamens A, Sivakumar R, Kato T, Yun D-W, Santhoshkumar C, Narayanan KK, Peacock JW, Dennis ES (2002) An *iAc/Ds* gene and enhancer trapping system for insertional mutagenesis in rice. Funct Plant Biol 29:547–559

Vorst O, Kock P, Lever A, Weterings B, Welsbeek P, Smeekens S (1993) The promoter of the *Arabidopsis thaliana* plastocyanin gene contains a far upstream enhancer-like element involved in chloroplast-dependent expression. Plant J 4:933–945

Walden R (2002) T-DNA tagging in a genomics era. Crit Rev Plant Sci 21:143–165

Webb KJ, Skot L, Nicholson MN, Jorgensen B, Mizen S (2000) *Mesorhizobium loti* increases root-specific expression of a calcium-binding protein homologue identified by promoter tagging in *Lotus japonicus*. Mol Plant Microbe Interact 13:606–616

Wei W, Twell D, Lindsey K (1997) A novel nucleic acid helicase gene identified by promoter trapping in *Arabidopsis*. Plant J 11:1307–1314

Wingender E, Chen X, Heh, R, Karas H, Liebich I, Matys V, Meinhardt T, Pruss M, Reuter I, Schacherer F (2000) TRANSFAC: an integrated system for gene expression regulation. Nucleic Acids Res 28:316–319

Wu C, Li X, Yuan W, Chen G, Kilian A, Li J, Xu C, Li X, Zhou DX, Wang S, Zhang Q (2003) Development of enhancer trap lines for functional analysis of the rice genome. Plant J 35:418–427

Yamagata H, Yonesu K, Hirata A, Aizono Y (2002) TGTCACA motif is a novel *cis*-regulatory enhancer element involved in fruit-specific expression of the *cucumisin* gene. J Biol Chem 277:11582–11590

Yamamoto YY, Tsuhara Y, Gohda K, Suzuki K, Matsui M (2003) Gene trapping of the *Arabidopsis* genome with a firefly luciferase reporter. Plant J 35:273–283

Yang TT, Cheng L, Kain SR (1996) Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein. Nucleic Acids Res 24:4592–4593

Yang W, Jefferson RA, Huttner E, Moore JM, Gagliano WB, Grossniklaus U (2005) An egg apparatus-specific enhancer of *Arabidopsis*, identified by enhancer detection. Plant Physiol 139:1421–1432

Yang Y, Li R, Qi M (2000) *In vivo* analysis of plant promoters and transcription factors by agroinfiltration of tobacco leaves. Plant J 22:543–551

Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldman KA, Meyerowitz EM (1990) The protein encoded by the *Arabidopsis* homeotic gene *Agamous* resembles transcription factors. Nature 346:35–39

Yoshida K, Shinmyo A (2000) Transgene expression systems in plants, a natural bioreactors. J Biosci Bioeng 90:353–362

Zheng MS, Takahashi H, Miyazaki A, Hamamoto H, Shah J, Yamaguchi I, Kusano T (2004) Up-regulation of *Arabidopsis thaliana NHL10* in the hypersensitive response to *Cucumber mosaic* virus infection and in senescing leaves is controlled by signaling pathways that differ in salicylate involvement. Planta 218:740–750

Chapter 24
Plant Genome Engineering Using Zinc Finger Nucleases

Sandeep Kumar and William F. Thompson

Abstract Gene targeting is a valuable tool to precisely integrate transgenes into predefined genomic locations. This technique allows transgenes to be targeted into any chosen locus in the genome to create plants with novel traits. Gene targeting by homologous recombination (HR) is also a useful approach for studying gene function because it allows the modification of specific genes, or sequences, within their native genomic context. A major aim of plant biotechnology now is the establishment of techniques for the precise manipulation of plant genomes. Until recently, despite multiple promising approaches, gene targeting in plants remained an elusive goal. Recently, however, zinc finger endonucleases (ZFN) have provided a means by which targeted double stranded breaks can be introduced at pre-defined genomic locations, and thus to stimulate recombination and repair processes at specific target sites. ZFN-mediated gene targeting is likely to become a powerful technique for precise modification and directed mutagenesis of plant and mammalian genomes.

24.1 Introduction

Although many agronomic traits may be manipulated through conventional breeding, plant transgenics will play a larger and larger role in future plant improvement programs for food, fiber, and novel uses such as bioenergy, chemical feedstocks, and pharmaceutical production. Plant transformation is also important in understanding the fundamental biology of plants. A sampling of the plant molecular biology literature in 2002 revealed that transgenic plants are used as an important research tool in about a half of the refereed publications (Miki and McHugh 2004).

Since the pioneer transformation efforts made in early 1980s, a significant advances have been made and it is now possible to transform several crop species across different genera. However the transgenes with the current generation of transformation technology are not only integrated randomly into the plant genome

S. Kumar (✉) and W.F. Thompson
Plant Gene Expression Laboratory, Department of Plant Biology, Campus Box 7550, NC State University,Raleigh NC, 27695
e-mail: skumar3@unity.ncsu.edu

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_24,

but frequently they are inserted in the form of complex multiple arrays of complete and partial transgenes. The complex integration patterns are associated with expression anomalies, including gene silencing. In addition to complex integration patterns, unpredictable transgene expression may also reflect characteristics of the chromatin at the site of transgene integration. The latter phenomenon is termed "chromosomal position effects" (Alberts and Sternglanz 1990; Dean et al. 1988) in which gene expression in certain regions in the genome is highly active whereas in other regions it is not (Allen et al. 1988; Mahy et al. 2002) reviewed in (Taddei et al. 2004). According to the idea of chromosomal position effects, there is a greater probability that the gene will be expressed at a high level if a transgene integrates into a region with a 'permissive' chromatin structure (such as already characterized by high transcriptional activity) (Mirkovitch et al. 1984). Conversely, if the transgene integrates into a transcriptionally inactive region with a repressive chromatin structure, there is a high probability that the transgene will be poorly expressed or even silenced.

The ability to insert transgenes into genomic sites with known expression characteristics would be a major step towards more efficient and robust transformation systems. Procedures to accomplish such specific insertions are collectively known as "gene targeting". Gene targeting can be obtained using site specific recombination (SSR) systems or via homologous recombination (HR). SSR requires a site-specific recombinase, which can catalyze the recombination between target recognition sites that are highly specific for the recombinase being used(Lyznik et al. 2003). Gene targeting using SSR is a two-step process. First, a target recognition site is pre-inserted into the genome and then donor transgene is inserted into the target site leading to gene targeting. This system has a limited application since recognition site needs to be placed before transgene can be targeted. In contrast, homologous recombination does not require exogenous recombinase but depends on host recombination systems to recognize homology between the donor sequence and endogenous sequences in the recipient genome. Gene targeting based on HR is a process in which a DNA molecule introduced into a cell replaces the corresponding chromosomal segment by homologous recombination, and thus presents a precise way to manipulate the genome (Porteus and Carroll 2005). Gene targeting is a valuable tool to precisely integrate transgene into predefined genomic location (Day et al. 2000). Transgenes can be targeted using HR to any chosen locus in the genome to create plants with novel traits (Kumar and Fladung 2001). In addition, HR techniques support "gene replacement" strategies that permit gene knockouts and structure/function studies of gene variants in their native genomic context (Jasin et al. 1996). A robust HR-based gene targeting system has been a longtime goal in plant genetic engineering.

24.2 Homologous Recombination and Gene Targeting in Plants

HR targeting requires significant homology (Hasty et al. 1991) between target and donor sequences and uses the endogenous recombination machinery of the plant system. In most eukaryotes, such as animal cells, HR targeting has been achieved

at frequencies efficient enough (1% or higher) for application in gene therapy (Jasin et al. 1996; Urnov et al. 2005). In contrast, HR targeting in plants remains inefficient and awaits major breakthrough to allow routine use (reviewed in Iida and Terada 2004; Puchta 2002, 2003). Initial gene targeting attempts using direct gene transfer (Halfter et al. 1992; Paszkowski et al. 1988) or *Agrobacterium*-mediated T-DNA transformation (Lee et al. 1990; Offringa et al. 1990) were directed at the restoration of function for previously inserted defective selectable marker genes to estimate the frequency of HR-mediated gene targeting events. Regardless of the plant species (tobacco or Arabidopsis) or the transformation method used, targeting frequencies in all these studies were low (10^{-4} to 10^{-5}). Given the successful gene targeting in mouse, efforts were made to improve gene targeting in plants by using longer regions of homology in the transferred DNA (Thykjaer et al. 1997) or keeping negative selectable markers outside the homology region of the targeting vector (Gallego et al. 1999; Risseeuw et al. 1997; Vergunst and Hooykaas 1999). However, neither of these techniques resulted in a significant increase in targeting frequencies.

Some of the initial reports on gene targeting in plants have unfortunately not been reproducible. For example, Eric Lam's group reported targeting of the *TGA3* locus in Arabidopsis and obtained one targeted callus out of 2,580 tissue culture transformation events. The targeted callus was found to be chimeric and no regeneration was obtained to confirm the gene targeting (Miao and Lam 1995). Similarly Martin Yanofsky's group attempted to target an Arabidopsis *AGL5* MADS-box gene and obtained one knockout mutant out of 750 transgenic plants (Kempin et al. 1997). However, these findings have not been reproduced, and the results appear to be controversial (Liljegren and Yanofsky 1998; Puchta 1998).

Terada et al. (2002), working with rice, were first to describe seemingly reproducible gene targeting system (Iida and Terada 2004, 2005; Terada et al. 2002, 2004). Their approach relies mainly on using powerful negative selection to eliminate random integration events. Using this procedure they modified *Adh2* locus in rice with a frequency of approximately 2% of surviving callus (Reviewed in Terada et al. 2007). This method does not increase the frequency of gene targeting per se, but is rather designed to increase the efficiency with which rare targeting events can be recovered, using a stringent combination of positive and negative selection. The technique therefore requires very efficient tissue culture and regeneration procedure.

It may be possible to enhance homologous recombination by overexpressing recombination enzymes in plant cells. Avraham Levy's group recently utilized *RAD54*-mediated chromatin remodeling to enhance gene targeting in plants (Shaked et al. 2005). They reported that expression of the yeast *RAD54* gene enhances the gene-targeting frequency in *Arabidopsis* by an average of 27-fold. *RAD54* is a member of the *SWI2 SNF2* superfamily of chromatin remodeling genes defined by the presence of conserved ATPase and helicase motifs in the proteins. This gene promotes strand invasion, an essential step in recombination between homologous DNA segments in yeast (Tan et al. 2003). This work provides a significant advance toward the goal of precision gene modification in plants (Puchta and Hohn 2005). While no noticeable side effects of overexpressing *RAD54* were reported, more exhaustive analysis for several generations is needed before a definitive conclusion is drawn.

The major breakthrough in plant gene targeting field however comes from a recent report from D. Voytas's group, who used Zinc Finger Nucleases (ZFN) to direct the creation of double-strand breaks and therefore increase the frequency of gene targeting at specified locations (Wright et al. 2005). In this chapter we will focus on the concept and potential application of this emerging gene targeting tool for plant systems.

24.3 Double-Strand Breaks at the Target Site Stimulate Homologous Recombination

Transgene integration is a part of genomic break and repair process in living organisms. Genomic double strand breaks (DSBs) are the key intermediates in these reactions. Once the DSBs are created, the cell's recombination machinery repairs them using HR or nonhomologous end joining (NHEJ). The process of HR uses an undamaged homologous segment of DNA as a template from which to copy the information across the break (Fig. 24.1). The homologous segment is usually a sister chromatid, but can also be an artificially-supplied donor DNA molecule. Because it recovers a normal copy of the damaged DNA, HR is the most accurate form of DSB repair. An alternative pathway of DSB repair is NHEJ, which joins ends without regard for homology and often results in small, localized deletions and/or insertions (Porteus and Carroll 2005).

Genomic DSBs play an important role in homologous recombination in eukaryotes. For example, factors such as X-rays (Lebel et al. 1993) or methyl methanesulfonate (Puchta et al. 1995) that induce random DSBs have been shown to enhance homologous recombination in plants. Among the strategies to engineer a given genetic locus, the use of rare cutting DNA endonucleases such as meganucleases has emerged as a powerful tool to increase homologous gene targeting through the generation of a DNA double strand break (DSB). Meganucleases recognize large (>12 bp) sequences, and can therefore cleave their cognate site without affecting global genome integrity (Arnould et al. 2007). Yeast mitochondrial homing endonuclease *I-Sce*I (Perrin et al. 1993), a natural meganuclease with 18-bp recognition site has been used to induce genomic DSBs and their repair by homologous recombination in mouse cells (Rouet et al. 1994; Choulika et al. 1995). Similarly in plants, genomic DSBs induced by *I-Sce*I can be repaired by HR using exogenously supplied donor DNA (Kumar and Fladung 2002). Puchta and colleagues tested this system by creating a transgenic locus containing an *I-Sce*I site into the tobacco genome. Transient expression of the *I-Sce*I gene then created a DSB at the inserted target site, resulting in a 100-fold increase in homologous integration frequency at that site (Puchta et al. 1996). This study provides an elegant proof-of-concept for enhanced HR at DSBs.

The induction of a DSB is therefore one of the rate-limiting steps in gene targeting in plants. Thus a gene targeting technique based on induction of DSBs at pre-defined genomic sites would be useful for plants. The homing endonucleases do not provide any flexibility in the choice of genomic regions for targeting as they

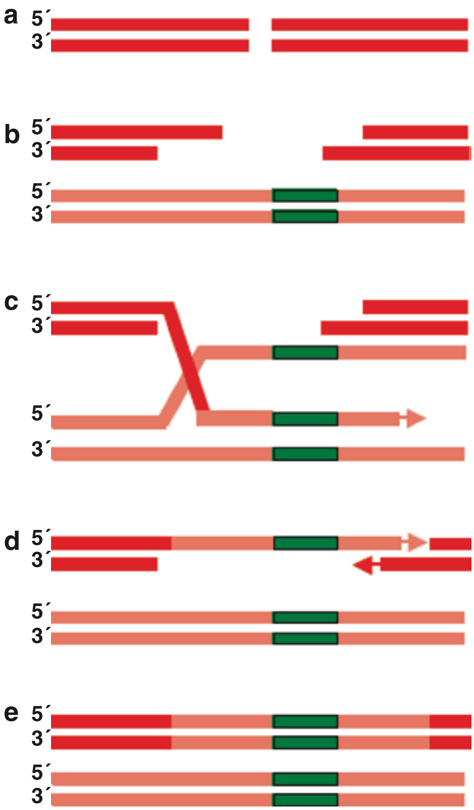


Fig. 24.1 Homology-directed DSB repair and gene targeting in plant cells. (**a**) DSB in the genomic target recipient site, which can occur spontaneously or be created specifically using ZFNs. (**b**) The DSB is processed to form free 3′ single strand tails. An extra-chromosomal homologous donor containing a functional transgene (*in green*) is also shown. (**c**) To repair the DSB, the free 3′ end invades the homologous donor to form a D-loop. After strand invasion, primed DNA synthesis occurs to generate new undamaged DNA using the donor DNA strand as a template. (**d**) The repair process is completed by the annealing of the new strand of DNA with its original partner and subsequent gap filling DNA synthesis. (**e**) The result is a modification of the locus, which now includes the transgene. The donor sequence remains unchanged; the recombination process is called gene conversion (reprinted from Kumar et al. (2006) with permission from Elsevier.)

require a prior transgenic step to introduce the cleavage site into the plant genome. This previous integration event cannot be directed, so targeting a specific region of the genome would require screening large numbers of random events. Custom-made nucleases designed to bind and cleave a specific DNA sequence could be ideal approach for obtaining unique DSBs at pre-determined sites. Several attempts have been made to modify DNA binding domains of natural meganucleases to obtain new tailored cleavage sites (see Arnould et al. 2007), but this approach has not yet been successful.

24.4 Zinc Finger Nucleases

Zinc finger nucleases (ZFN) provide an attractive tool for improving the efficiency of gene targeting because they can be designed to recognize naturally any DNA sequence. By inducing DSBs at pre-defined locations, they can stimulate the cell's endogenous HR machinery to incorporate exogenous DNA by homologous recombination. The zinc finger is a compact protein module that has a remarkably simple mode of interaction with DNA, making it ideal for protein engineering (Choo and Isalan 2000). ZFN are artificial fusion proteins that link a zinc finger DNA binding domain to a nonspecific nuclease domain. Chandrasegaran's lab was the first to create these artificial nucleases by fusing the nonsequence-specific cleavage domain of the type II restriction endonuclease *Fok*I (Fn domain) to a zinc finger DNA binding domain (Fig. 24.2) (Kim et al. 1996; Kim and Chandrasegaran 1994). The modular nature zinc finger proteins offer an attractive framework for designing nucleases (ZFN) with tailor-made sequence-specificities (Bibikova et al. 2001, 2003; Kandavelou et al. 2005).

The zinc fingers can be artificially designed, which provide cleavage specificity and can be modified to recognize essentially any genomic site. The DNA-binding domain of a ZFN comprises a string (usually 3 or 4) of 'zinc finger motifs', each of which is a stretch of around 30 amino acids, stabilized by a zinc ion, that binds to a particular three base DNA sequence (Durai et al. 2005). The zinc finger motif varies according to the DNA sequence to which it binds, and the motif can be changed while maintaining the remaining amino acids as a consensus backbone to generate ZFN with different sequence specificities (Mani et al. 2005). Generally, three (or four) such zinc finger domains are linked together in tandem to generate a zinc finger protein that binds to a 9 (or12)-bp site, which is a composite of the individual DNA triplet subsites recognized by each of the three zinc finger motifs.

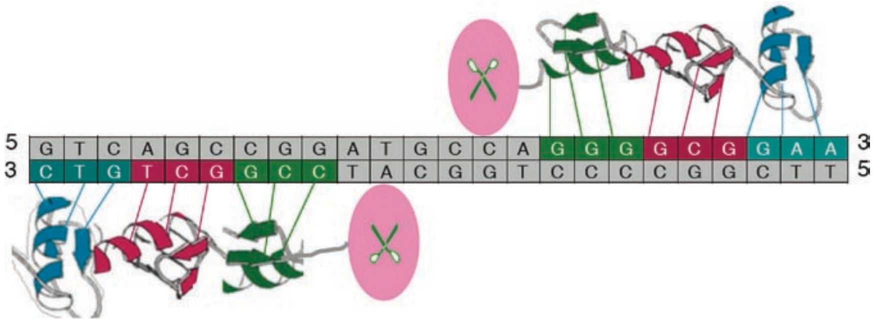


Fig. 24.2 Schematic representation of a three-finger ZFN. Zinc fingers that recognize and bind nucleotide triplets of the target sequence are depicted. Zinc finger domains are fused to the cleavage domain of endonuclease *Fok*I (*shown as scissors*) to create the ZFN. The three-finger ZFN requires two adjacent 9 bp recognition sites (*shown in color*) in an inverted orientation in order to dimerize and produce a DSB at the target sequence (modified from Kumar et al. (2006) with permission from Elsevier)

The linked zinc finger motifs then are fused to the nonspecific endonuclease domain of *Fok*I restriction enzyme (Fig. 24.2). Thus, once the zinc finger dimerization bound to the target site the nuclease domain can introduce a site-specific DSB into the genome. The strategy requires binding of a pair of 3 (or 4) finger ZFN monomers each recognizing a 9 (or 12)-bp inverted site because dimerization of the *Fok*I cleavage domain is required to produce a DSB. Therefore, 3 (or 4)-finger ZFN effectively have an 18 (or 24)-bp recognition sites, which is long enough to specify a unique address within the genome (Kandavelou et al. 2005).

Bibikova et al. extensively studied ZFN-mediated cleavage and recombination in *Xenopus laevis* oocyte system (Bibikova et al. 2001). They reported that the most efficient cleavage was obtained when the binding sites were inversely oriented and separated by six nucleotides, and when there was no intentional linker between the zinc finger and nuclease domains. Porteus and Baltimore demonstrated that ZFN could stimulate gene targeting in mammalian somatic cells (Porteus and Baltimore 2003). In these experiments, a mutated green fluorescent protein (GFP) reporter gene containing recognition sites for known zinc finger DNA-binding domains was integrated as a single copy into the genome of the human embryonic kidney cell line. The gene targeting was measured by the correction of the mutant GFP target gene by a transfected donor plasmid. Using four finger ZFN Urnov et al. (2005) recently reported that appropriately designed ZFN can cleave an endogenous human gene in cultured cells and lead to targeted gene replacement in up to 20% of the cells. The target was the gene for interleukin 2 receptor γ (IL2Rγ), a cytokine receptor that is required for T-cell development and the establishment of a functional immune system. This HR-mediated modification frequency is remarkably high compared to the rate of 0.001% generally obtained when no DSBs are introduced.

24.5 Zfn in Plant Systems

The first report on ZFN in plants was from Drew's lab (Lloyd et al. 2005), who showed that ZFN can cleave and stimulate mutations at specific genomic sites in Arabidopsis. They obtained targeted mutagenesis in *Arabidopsis* using a transgene containing both a ZFN gene driven by a heat-shock promoter, and its target. Targeted mutations, as high as 0.2 per target, were obtained demonstrating that ZFN can be highly efficient for targeted mutagenesis of plant genes. This work provided a very good platform to extend the use of ZFN to establish homology-directed gene targeting system for plants.

D. Voytas's group recently reported successful ZFN-mediated gene targeting in plant cells (Fig. 24.3) (Wright et al. 2005). To measure the HR and ZFN-mediated target modification they designed an artificial target gene containing a translational fusion between b-glucuronidase (GUS) and neomycin phosphotransferase (NPTII) that conferred both selectable and screenable phenotypes to plant cells. The target gene, *GUS:NPTII*, was made non-functional by deleting 600 bases that encode the active site of GUS and part of the ATP-binding domain of NPTII. A cleavage site

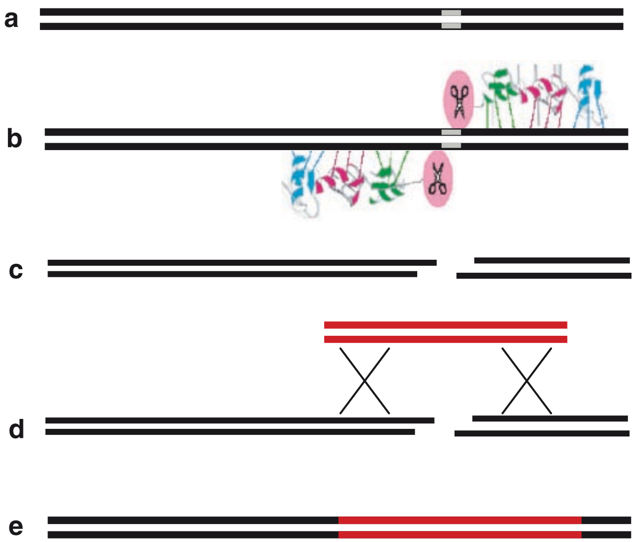


Fig. 24.3 Schematic representation of ZFN-mediated gene targeting. (**a**) Genomic sequence containing recognition site shown in grey. (**b**) ZFN binds the target sequence and (**c**) introduces a DSB into the target site. (**d**) The break is then repaired via homologous recombination (as shown in Fig. 24.1) using a fragment of donor DNA (*shown in red*) carrying the desired sequence. (**e**) The process leads to precise modification of the target gene, which now contains the desired donor sequence

specific to the synthetic ZFN was inserted at the site of the deletion in the target gene. The donor DNA molecule for homologous repair contained the 600 bases missing from the target gene, but lacked the 457 bp of the 5′ end of *GUS* coding sequence so that it would be non-functional when integrated randomly into the genome. Only homologous integration of the donor into the defective target would restore *GUS:NPTII* function.

Protoplasts isolated from ten tobacco plants containing the defective target gene were electroporated with a mixture of donor DNA for gene targeting and DNA encoding the ZFN. Using this test system, Wright et al. obtained HR-based gene targeting in 10% of the transformed protoplasts, a frequency which is comparable to earlier work in plants where *I-Sce*I was used to create DSBs (Puchta et al. 1996). This is a 10^4- to 10^5-fold enhancement over frequencies previously reported with gene targeting procedures without exogenously expressed nuclease or recombinase gene (reviewed in Puchta 2002). One in five characterized gene targeting events was free of DNA insertions or deletions sustained during repair of the target locus. Hence the frequency of ‘pure’ HR events is one per 50 transformation events.

Although it is well documented in the literature that creating DSB can stimulate HR in plants, tools to create DSB into a desired genomic site were previously not available. ZFN has emerged as a novel tool to make a unique break in potentially any genomic site, in order to stimulate recombination and gene targeting at that site.

24.6 Conclusion and Future Prospects

Although these recent reports open the door to a new era of genetic engineering, gene targeting in plants continues to be far from routine. Several challenges remain before such applications can be realized and ZFN approaches become widely adopted for gene targeting in plants First, applicability of the ZFN approach needs to be tested in wide variety of plant species and cultivars. Secondly, the method for delivering ZFN and repair substrate to cells must be optimized. Thirdly, our understanding of the process of homologous recombination itself needs to be improved (Porteus and Carroll 2005).

Efforts to broaden the applicability of the approach will require design of ZFN targeting a greater variety of gene targets. ZFN have so far been tested in Arabidopsis plants and tobacco protoplasts. More challenging perhaps would be to adapt this technique to broad range of plant species or cultivars that remain recalcitrant to transformation.

The optimization of ZFN design must address both specificity of the enzyme and potential cytotoxicity to the host plant. The genomic integrity and stability of treated cells should be maintained except the desired break at the targeted chromosomal site (Kumar et al. 2006). Addressing these issues will necessitate thorough analysis of the mechanisms of gene target recognition and binding for the zinc finger component. Potential cytotoxicity may be addressed by adding more zinc finger motifs to create greater sequence specificity and (Durai et al. 2005) and by transient or regulated expression of ZFN in the target cell population (Kandavelou et al. 2005).

A final challenge will be to increase our understanding of the process of homologous recombination itself. The development of gene targeting in plants appears to be a fusion of two independent but synergistic activities: the ZFN-mediated targeted DSBs, and the repair process with exogenously supplied homologous donor sequence. It should be noted that though ZFN can create targeted breaks, it is the cell's DNA repair mechanisms that are engaged to repair the break via non homologous end joining (NHEJ) or via HR. Chromatin structure may also constrain HR by making the target locus inaccessible to the ZFN, or to the donor DNA, thus potentially restricting the choice of target sites. Hence targeted breaks combined with chromatin remodeling (Shaked et al. 2005), or manipulation of genes related to HR and NHEJ would further facilitate the process of homologous recombination and gene targeting in plants.

Acknowledgement This work was funded by grants from National Science Foundation (DBI-0110023 and DBI-0421651). We thank Dr. George Allen and Dr. Steve Spiker for their helpful encouragement and discussions.

References

Alberts B, Sternglanz R (1990) Chromatin contract to silence. Nature 344:193–194

Allen ND, Cran DG, Barton SC et al. (1988) Transgenes as probes for active chromosomal domains in mouse development. Nature 333:852–855

Arnould S, Perez C, Cabaniols JP et al. (2007) Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. J Mol Biol 371:49–65
Bibikova M, Carroll D, Segal DJ et al. (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. Mol Cell Biol 21:289–297
Bibikova M, Beumer K, Trautman JK et al (2003) Enchancing gene targeting with designed zinc finger nucleases. Science 300:764
Choo Y, Isalan M (2000) Advances in zinc finger engineering. Curr Opin Struct Biol 10:411–416
Choulika A, Perrin A, Dujon B, Nicolas JF (1995) Induction of homologous recombination in mammalian chromosomes by using the I-*SceI* system of *Saccharomyces cerevisiae*. Mol Cell Biol 15:1968–1973
Day CD, Lee E, Kobayashi J et al. (2000) Transgene integration into the same chromosome location can produce alleles that express at a predictable level, or alleles that are differentially silenced. Genes Dev 14:2869–2880
Dean C, Jones J, Favreau M et al. (1988) Influence of flanking sequences on variability in expression levels of an introduced gene in transgenic tobacco plants. Nucleic Acids Res 16:9267–9283
Durai S, Mani M, Kandavelou K et al. (2005) Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. Nucleic Acids Res 33:5978–5990
Gallego ME, SirandPugnet P, White CI (1999) Positive-negative selection and T-DNA stability in Arabidopsis transformation. Plant Mol Biol 39:83–93
Halfter U, Morris PC, Willmitzer L (1992) Gene targeting in *Arabidopsis thaliana*. Mol Gen Genet 231:186–193
Hasty P, Rivera-Perez J, Bradley A (1991) The length of homology required for gene targeting in embryonic stem cells. Mol Cell Biol 11:5586–5591
Iida S, Terada R (2004) Tale of two integrations, transgene and T-DNA: gene targeting by homologous recombination in rice. Curr Opin Biotechnol 15:132–138
Iida S, Terada R (2005) Modification of endogenous natural genes by gene targeting in rice and other higher plants. Plant Mol Biol 59:205–219
Jasin M, Moynahan ME, Richardson C (1996) Targeted transgenesis. Proc Natl Acad Sci USA 93:8804–8808
Kandavelou K, Mani M, Durai S et al. (2005) “Magic” scissors for genome surgery. Nat Biotechnol 23:686–687
Kempin SA, Liljegren SJ, Block LM et al. (1997) Targeted disruption in Arabidopsis. Nature 389:802–803
Kim YG, Chandrasegaran S (1994) Chimeric restriction endonuclease. Proc Natl Acad Sci USA 91:883–887
Kim YG, Cha J, Chandrasegaran S (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A 93:1156–1160
Kumar S, Allen GC, Thompson WF (2006) Gene targeting in plants: fingers on the move. Trends Plant Sci 11:159–161
Kumar S, Fladung M (2001) Controlling transgene integration in plants. Trends Plant Sci 6:155–159
Kumar S, Fladung M (2002) Gene Targeting in plants In: Jain SM, Brar DS, Ahloowalia BS (eds) Molecular techniques in crop improvement. Kluwer Academic, Dordrecht, pp 481–500
Kumar S, Allen GC, Thompson WF (2006) Gene targeting in plants: finger on the move.Trends Plant Sci 11:159–161
Lee KY, Lund P, Lowe K et al. (1990) Homologous recombination in plant cells after Agrobacterium-mediated transformation. Plant Cell 2:415–425
Liljegren SJ, Yanofsky MF (1998) Towards targeted transformation in plants… response: targeting Arabidopsis. Trends Plant Sci 3:79–80
Lebel EG, Masson J, Bogucki A, Paszkowski J (1993) Stress-Induced Intrachromosomal Recombination in Plant Somatic Cells. Proc Natl Acad Sci U S A 90:422–426
Lloyd A, Plaisier CL, Carroll D et al. (2005) Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. Proc Natl Acad Sci U S A 102:2232–2237

Lyznik LA, Gordon-Kamm WJ, Tao Y (2003) Site-specific recombination for genetic engineering in plants. Plant Cell Rep 21:925–932
Mahy NL, Perry PE, Gilchrist S et al. (2002) Spatial organization of active and inactive genes and noncoding DNA within chromosome territories. J Cell Biol 157:579–589
Mani M, Kandavelou K, Dy FJ et al. (2005) Design, engineering, and characterization of zinc finger nucleases. Biochem Biophys Res Commun 335:447–457
Miao ZH, Lam E (1995) Targeted disruption of the TGA3 locus in *Arabidopsis thaliana*. Plant J 7:359–365
Miki B, McHugh S (2004) Selectable marker genes in transgenic plants: applications, alternatives and biosafety. J Biotechnol 107:193–232
Mirkovitch J, Mirault ME, Laemmli UK (1984) Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. Cell 39:223–232
Offringa R, de Groot MJA, Haagsman HJ, Does MP, van den Elzen PJM, Hooykaas PJJ (1990) Extrachromosomal homologous recombination and gene targeting in plant cells after Agrobacterium mediated transformation. EMBO J 9:3077–3084
Perrin A, Buckle M, Dujon B (1993) Asymmetrical recognition and activity of the I-SceI endonuclease on its site and on intron-exon junctions. EMBO J 12:2939–2947
Paszkowski J, Baur M, Bogucki A, Potrykus I (1988) Gene targeting in plants. EMBO J 7:4021–4026
Porteus MH, Baltimore D (2003) Chimeric nucleases stimulate gene targeting in human cells. Science 300:763
Porteus MH, Carroll D (2005) Gene targeting using zinc finger nucleases. Nat Biotechnol 23:967–973
Puchta H (1998) Towards targeted transformation in plants. Trends Plant Sci 3:77–78
Puchta H (2002) Gene replacement by homologous recombination in plants. Plant Mol Biol 48:173–182
Puchta H (2003) Towards the ideal GMP: homologous recombination and marker gene excision. J Plant Physiol 160:743–754
Puchta H, Hohn B (2005) Green light for gene targeting in plants. PNAS 102:11961–11962
Puchta H, Dujon B, Hohn B (1996) Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. Proc Natl Acad Sci U S A 93:5055–5560
Puchta H, Swoboda P, Hohn B (1995) Induction of intrachromosomal homologous recombination in whole plants. Plant J 7:203–210
Risseeuw E, Franke-van Dijk MEI, Hooykaas PJJ (1997) Gene targeting and instability of Agrobacterium T-DNA loci in the plant genome. Plant J 11:717–728
Rouet P, Smih F, Jasin M (1994) Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. Proc Natl Acad Sci U S A 91:6064–6068
Shaked H, Melamed-Bessudo C, Levy AA (2005) High-frequency gene targeting in Arabidopsis plants expressing the yeast RAD54 gene. PNAS 102:12265–12269
Taddei A, Hediger F, Neumann FR, Gasser SM (2004) THE FUNCTION OF NUCLEAR ARCHITECTURE: A Genetic Approach. Ann Rev Genet 38:305–345
Tan TL, Kanaar R, Wyman C (2003) Rad54, a Jack of all trades in homologous recombination. DNA Repair (Amst) 2:787–794
Terada R, Urawa H, Inagaki Y, Tsugane K, Iida S (2002) Efficient gene targeting by homologous recombination in rice. Nat Biotechnol 20:1030–1034
Terada R, Asao H, Iida S (2004) A large-scale Agrobacterium-mediated transformation procedure with a strong positive-negative selection for gene targeting in rice (*Oryza sativa* L.). Plant Cell Rep 22:653–659
Terada R, Johzuka-Hisatomi Y, Saitoh M, Asao H, Iida S (2007) Gene targeting by homologous recombination as a biotechnological tool for rice functional genomics. Plant Physiol 144:846–856
Thykjaer T, Finnemann J, Schauser L, Christensen L, Poulsen C, Stougaard J (1997) Gene targeting approaches using positive-negative selection and large flanking regions. Plant Mol Biol 35:523–530

Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S, Jamieson AC, Porteus MH, Gregory PD, Holmes MC (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature 435:646–651
Vergunst AC, Hooykaas PJJ (1999) Recombination in the plant genome and its application in biotechnology. Crit Rev Plant Sci 18:1–31
Wright DA, Townsend JA, Winfrey RJ, Jr., Irwin PA, Rajagopal J, Lonosky PM, Hall BD, Jondle MD, Voytas DF (2005) High-frequency homologous recombination in plants mediated by zinc-finger nucleases. Plant J 44:693–705

Chapter 25
Cisgenesis

Next Step in Classical Plant Breeding

Evert Jacobsen and Henk J. Schouten

Abstract Plant breeding is an ongoing activity that started many years ago through domestication of crops by farmers' selection. After discovery of Mendel's Law plant breeding became gradually more science driven. Nowadays, plant breeding is developing very rapidly because of the development of many new technologies and scientific disciplines that can be applied. Approaches for genetic modification (GM) of plants developed quickly in the eighties and nineties of last century, but it is the first technology that has not been widely accepted in the world by NGOs and consumers. GMO-regulations have been developed which are by the strict application obstructing the development of GM-varieties, especially in Europe. These regulations are based on the modification process and on transgenes originating from non-crossable species. These transgenes are a new gene pool for plant breeding. However, it turns out that cisgenes, which are genes from the plant itself or from crossable species, will be more and more available. They belong to the existing breeder's gene pool but they are treated in the regulation like transgenes. It is recommended to exempt from the regulation GM-plants that contain cisgenes only. This chapter provides a historical context of cisgenesis. Further, it discusses breeding approaches of autogamous, allogamous and vegetatively propagated crops. Options for cisgenesis in these kind op crops are presented. Some examples are disease resistance in potato and apple using R- and Avr-genes, hybrid seed production using genes for male sterility, or S_RNase genes for changing self-incompatibility. We regard cisgenesis as next important step in introgression breeding, using natural genes spatie. Cisgenesis has also been compared with intragenics and induced mutation breeding. We recommend less stringent oversight for intragenic plants, compared to transgenic plants especially when it concerns RNAi.

E. Jacobsen (✉) and H.J. Schouten
Plant Breeding, Wageningen University and Research Center, Wageningen, The Netherland
e-mail: evert.jacobsen@wur.nl

E. Jacobsen
Transforum Agribusiness & Rural Areas, Louis, Pasteurlaan 6, 2700 AB Zoetermeer, The Netherlands

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_25,

25.1 Introduction

25.1.1 Domestication of Crops and Traits

25.1.1.1 Initiation of Domestication

Domestication of the different crops has been a long-term process that started for our main staple food crops more than 10,000 years ago in the Middle East, Central America and South-East Asia. Initially, plant parts like fruits, seeds, tubers were collected in nature only. Wild plants around settlements were more frequently used. Some plant species became permanent residents of human settlements (camp followers) and a permanent food source. During a following phase growing conditions of camp followers were improved, as mankind prepared the soil and collected seeds for sowing. Selection of better individual plants started. Seed of domesticated material was collected and stored for next growing season. Many crop species gradually developed by selection for desired traits. During all development stages of agriculture from collection into modern cultivation, including fruit, flower and vegetable production in the field or in glasshouses, selection and adaptation to the new circumstances has been most crucial. Traits of importance for domestication were yield, flowering time, larger seeds or tubers, shattering resistance, absence of thorns or prickles and of bitter, poisonous compounds, uniformity, adaptation to modern agricultural practice, and many other traits. In some crops the genetic diversity is substantially reduced for alleles that have been particularly important for domestication. In maize that seemed to be the case for example for teosinte branched 1 gene controlling tillering and apical dominance (Buckler et al. 2001).

25.1.1.2 Farmer's Selection

Mankind has selected crop plants during thousands of years. This process of crop improvement was slow. The genetic aspects involved were poorly understood until the end of the 19th century. Farmers' selection to local circumstances and needs resulted in so-called landraces. Landraces are still the plant material for many farmers worldwide. An interesting aspect in the evolution of domestication is the detection of natural introgression by farmer selection in different agro-ecosystems. It means that in addition to introgression of genes by humans through interspecific crossing, selection, and back crossings, also unintended natural introgression has occurred. There are indications (Jarvis and Hudgkin 1999) that it occurs both in autogamous (self-fertilizing) and allogamous (cross fertilizing) crops. Also natural translocation has been observed (Schmidt 2002). There are strong indications that the phenomenon of translocation is occurring in wild plants as well as in crop plants (Lim et al. 1998). It means that human introgression as well as (induced) translocation procedures are not principally different from spontaneous processes occurring in nature. At present, farmers' selection and management of local crops is still very important.

This activity is imbedded in forms of traditional plant breeding. Over 40% of the world's agricultural area is still under management and seed supply of small-scale farmers (Jarvis and Hodgkin 1999).

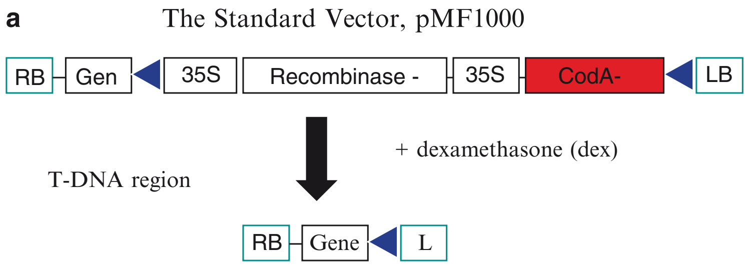
25.1.1.3 Professional Cross Breeding

Rediscovery of Mendel's laws at the end of the 19th century led to a shift to more science-based plant breeding. It stimulated professionalization of plant breeding and development of seed business. Combining useful traits by crossing and selection on the basis of classical genetics influenced crop improvement strongly, and led to more rapid selection for superior types. The impact of classical genetics on professional plant breeding started in the early twentieth century, but is still ongoing. It has stimulated domestication of crops in many ways.

Selection in plant breeding implies reduction of genetic variation. However, also widening of genetic variation is required for novel traits, such as resistances to biotic or a-biotic stress, further yield improvement, hybrid seed production and new quality traits. Additional genetic variation can be found in crossable relatives. The pool of genetic variation can be further extended by techniques such as embryo rescue, protoplast fusion, and recently genetic modification. In addition to domestication of crops by normal cross breeding, in which several traits from both crossing parents will be combined, it is also important to introgress specific or single traits from wild material into crops. Introgression and in some cases induced translocation breeding of specific traits simultaneously introduces so many unwanted alleles from the wild plant that pre-breeding is required, in order to remove the majority of these unwanted alleles. Unwanted alleles that are genetically tightly linked to the desired allele can bring serious problems. The problem of linkage drag with negative traits is frequently the main bottle neck in such approaches. Solving linkage drag problems may require many years of crossing and selection. Molecular markers can nowadays be helpful to reduce linkage drag problems or to speed up the solution, but it never completely removes the linkage drag. An example of linkage drag is insect resistance in lettuce that was closely linked to compact growth and rapid aging. Jansen (1997) described how Marker Assisted Breeding (MAS) was helpful to solve this linkage problem.

25.1.1.4 Cisgenic Breeding

A sometimes more efficient solution for introgression of alleles from wild germplasm into crops is cisgenesis (Schouten et al. 2006a). In this approach, the allele from the wild relative is isolated molecularly from the genomic DNA of that wild relative. Subsequently, this allele is transferred to the recipient crop plant through marker free genetic modification techniques (Fig. 25.1). Isolation of the allele from the donor plant is crucial here, as it is not flanked by undesired alleles anymore. By definition, cisgenesis prevents linkage drag. This is the major advantage of cisgenesis.



b *A.tumefaciens* without selection genes

c Mixed inoculation followed by genetic segregation

Fig. 25.1 Marker-free transformants can be obtained in three different ways (**a**) Normal transformation on selection medium followed by induced excision of the marker gene and helper genes using e.g. the standard vector pMF 1,000. The diagram is depicting a right and left border (RB, LB), the cisgene of interest, two recombination sites (*blue triangles*), and between these recombination sites the selection gene NptII and a recombinase construct, both controlled by 35S promoters. After transformation and selection, the whole segment between the recombination sites is removed on recombinase-mediated excision induced by dexamethasone (Dr. F.A. Krens, Plant Research International, Wageningen). (**b**) A marker-free *A. tumefaciens* strain with only agricultural gene(s) in the TDNA is used during transformation. Transformants are selected among many regenerants by PCR. (**c**) Co-inoculation with two *A. tumefaciens* strains containing the selection marker or the agricultural cisgene(s), respectively. The marker-free transformants can be selected among the segregants after a sexual step by selfing or back crossing with a wild type plant

Further, by definition, the cisgene contains its native introns and is preceded by its native promoter and followed by its native terminator (Schouten et al. 2006a, b). The cisgenic insertion contains only one or a few cisgenes and is not surrounded by other genes of the donor plant, nor by genes from a vector's backbone. Here domestication of the cisgene is obtained by surgical precise isolation of desired alleles from donor plants, whereas in introgression breeding many times uncontrolled meiotic recombinations of linked genes are required for removal of negative side effects.

Cisgenesis requires methods of genetic modification that do not leave behind transgenes for selection of transformed cells, for example coding for herbicide tolerance or for resistance to antibiotics. As shown in Fig. 25.1, nowadays different methods are applied to obtain such marker-free plants (Fig. 25.1). The first possibility is the removal afterwards of the selection genes and other helper genes by recombinase-based excision (Schaart et al. 2004). The second approach is transformation without a selection gene, but screening by means of PCR for transgenic regenerants containing the insert (Vetten de et al. 2003). The third approach is infection with two *Agrobacterum* strains, the first one containing the selection marker and the second one the agricultural cisgene(s) of interest (Yu et al. 2006). After sexual crossing with wild type plants or after selfing, cisgenic plants can be

selected from the segregating offspring. Plants with only the cisgene(s) are selected and the transgenic plants with only the selection gene or with both selection gene and the cisgene(s) are discarded.

25.2 Intragenics, RNAI and Induced Mutation Breeding

Cisgenesis is very restricted regarding the allowed gene sequences, keeping as close to the classical breeding as possible. Another approach is intragenics. In intragenics, different DNA fragments from the breeder's gene pool are put together into new combinations, such as the coding sequence from one gene, and a promoter from another gene, and are inserted by means of genetic modification techniques into the recipient plant. By making novel combinations of native coding sequences and native regulatory sequences, the expression pattern of genes can be changed. In this sense intragenics is less strict and not as close to classical breeding as cisgenesis. In case of intragenics, all tDNA is derived from DNA from crossable relatives, even the border sequences. This approach has been described by Rommens (2004) and Rommens et al. (2007). Intragenics is also based on a marker-free approach.

Silencing by means of RNAi can also be obtained by means of intragenics, in case the whole tDNA sequence is derived from plant DNA. RNAi is a very important approach mimicking natural or induced loss of function mutations in plants.

A major difference between induced mutation breeding and the RNAi approach is that loss of function by means of a mutation is inherited recessively, but RNAi inherits dominantly and can therefore also be used in polyploid and vegetatively propagated crops. The example of simultaneously improving storage and processing characters in processing potatoes is clear (Rommens et al. 2006).

25.3 Regulation of Cisgenesis and Intragenics

As cisgenic plants are very similar to traditional plants, and at least as safe as traditionally bred plants or plants from (induced) mutation breeding or induced translocation breeding, Schouten et al. (2006a, b) and Jacobsen and Schouten (2007) have proposed to exempt cisgenesis of plants from the regulation on deliberate release of GMOs into the environment, therewith clearing cisgenic plants in a timely and cost-effective manner. Rommens et al. (2007) proposed this also for intragenic plants. The main reason is the source of the genetic material, which is within the species or within sexual compatible species, available to conventional plant breeding. The main difference between cisgenesis and intragenics is that cisgenesis used natural genes including their native promoter, whereas intragenics allows also novel combinations of promoters and coding sequences, present in the plant and which are or can be used in traditional plant breeding.

Table 25.1 Proposed changes for notification in GMO-regulations with respect of transgenes, cisgenes or intragenes in plants. The transformation process brings per definition a GM-plant. The DNA source determines the way regulation should be applied. Transgenes are partly from non crossable species, but intragenes and cisgenes are from the species itself or from crossable species. A. Baseline is natural complete genes. B. Baseline is natural functional sequences such as promotors, coding parts and terminators

		GMO regulation	
Categories	Type of genes	A	B
1	New transgenes	Full	Full
2	New events in existing gene-crop combination	Partial	Partial
	Intragenes	Partial	Exempted
3	Cisgenes	Exempted	Exempted

Table 25.1 shows a proposal for deregulation of cisgenic and intragenic plants. Both in cisgenesis and intragenics, the genes used belong to the gene pool of the conventional breeder. The conventional breeding is regarded as the baseline for the GMO regulation. As cisgenesis uses natural genes with their native promoters from the breeder's gene pool, it makes sense in treating cisgenic plants in a similar way as traditionally bred plants (Schouten et al. 2006a, b; Jacobsen and Schouten 2007). Intragenics uses the same gene pool too, and therefore should be deregulated too. However, expression patterns may be obtained that are very unlikely through conventional breeding. Intragenics can therefore result into plants that cannot be obtained by means of conventional breeding. Therefore we propose a partial deregulation for intragenics. RNAi is in this sense a special application of intragenics, mimicking natural or induced mutations. Also in this case deregulation is logic. It results in loss of function of genes that may also be obtained by means of induced mutations. In case RNAi is designed for a sequence that appears at more loci of the plant's genome, such as a conserved domain of a gene family, it may interfere with a group of genes. This exceeds the effect of a mutation in one locus. Therefore we propose here partial deregulation. In Table 25.1 we show also the situation in which functional sequences as promotors, coding parts and terminators of intragenes are seen as smallest natural units for exemption. In this case intragenes as well as cisgenes would be exempted.

25.4 Autogamous Crops

This group of plants contains many important annual crops such as cereals, sweet pepper, tomato, pulses and lettuce. They can be subdivided into real diploids like tomato, barley, rice, pea, soybean and lettuce and more complex disomic allopoly-ploids like wheat and oat. We will restrict this part to diploid autogamous crops. The genetic variation in these crops is mainly found within the species, varying from other varieties to wild material or to a restricted number of related species. All these mentioned crops are advanced regarding their breeding development.

25.4.1 Introgression and Pre-Breeding

Improved varieties have to combine more and more traits not only originating from their indigenous germplasm within the species and existing varieties but for an increasing number of traits, like disease resistance, quality traits and male sterility, also from crossable wild relatives. In practice this is introgression breeding by interspecific hybridization followed by back crosses with the cultivated plant under simultaneous selection for the trait of interest as shown in Fig. 25.2. Bottleneck in this approach is frequently linkage drag with neighboring wild alleles flanking the allele of interest (Jansen 1997).

Because of the great number of co-introgressing unwanted alleles from the wild germplasm, pre-breeding is often performed. Pre-breeding activities are frequently focused on solving the problem that other negative traits are genetically linked to the gene of interest. MAS is a nice way to speed up this process of domesticating a trait, however, a lot of efforts are needed before it can be applied properly (William et al. 2007). The problem of linkage drag is increasing if simultaneously several traits have to be introduced from different wild plants, specifically from less related species and/or when these traits are linked on the same locus of the same chromosome in different backgrounds.

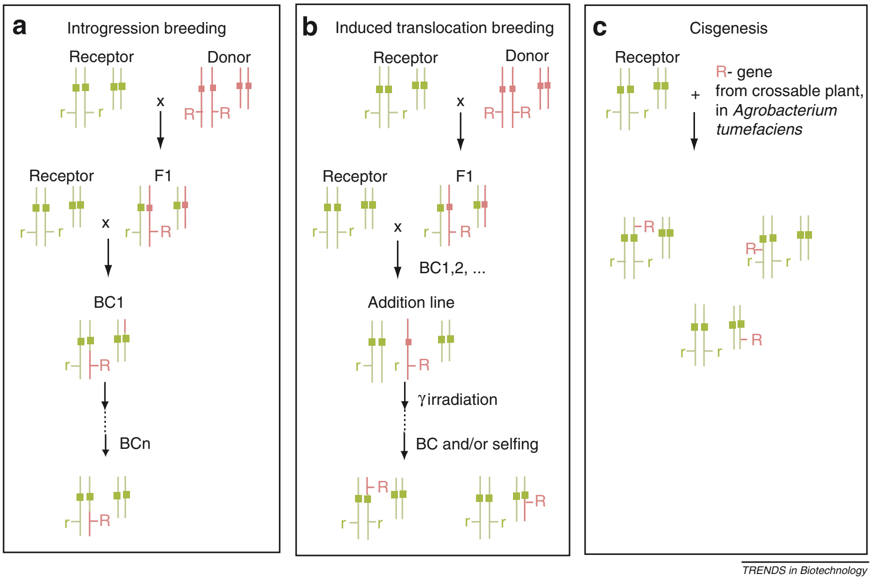


Fig. 25.2 The most important steps in introgression, induced translocation and cisgenic breeding with a resistance (R) gene. These result in (**a**) introgression, (**b**) induced translocation of the R-gene with linkage drag, and (**c**) cisgenic R-gene insertion without linkage drag (Jacobsen and Schouten 2007)

The availability of molecularly isolated genes and marker-free transformation methods would help breeding. For seed propagated crops co-inoculation with two *Agrobacterium tumefaciens* strains followed by genetic segregation of the selection marker and the agricultural traits, is a good option for obtaining marker-free transformation. It is known that T-DNA from both strains are frequently inserted at different chromosomes in the same transformed cell, so that they will segregate independently after crossing (Yu et al. 2006).

25.4.2 Induced Translocation Breeding

In allopolyploid crops, which are more complex, introgression breeding is not always possible. In examples like wheat, cross breeding is usually restricted to wild material from the crop species itself or to one of the parental species. If other species are needed, introgression breeding is frequently more complex. Main reason for that is the allopolyploid nature of this crop plant. The allopolyploid nature is connected to disomic inheritance of traits instead of polysomic inheritance in autopolyploid crops. In wheat disomic inheritance is occurring because of the *Ph*-gene (Friebe et al. 1996) which is suppressing interspecific chromosome pairing. The presence of this *Ph*-gene is not only suppressing interspecific chromosome pairing of the different genomes involved in the existing crop but also from other species during the process of introgression breeding. Suppression of chromosome pairing in interspecific hybrids of wheat, with species like *Aegilops umbellulata, Agropyron elongatum, Aegilops speltoides* (Jacobsen and Schouten 2007) is also obstructing considerably introgression of wild genes into the allopolyploid genome. This problem can be solved in different ways, such as inducing interspecific chromosome pairing by removing the *Ph*-gene temporally by nulli 5B-tetra 5A compensating genotypes or by selecting for natural suppression of this trait (Friebe et al. 1996). Fig. 25.2 shows induced translocation as another commonly used solution**.** Basic aspect is irradiation of monosomic addition lines that have been obtained by interspecific hybridization followed by back crosses with wheat under simultaneous selection for the novel trait of interest. The alien monosomic addition line is bearing the gene of interest. After irradiation of the monosomic addition line and selfing, resistant offspring is selected that lack the alien chromosome but is still possessing the functional alien resistance gene, translocated into the wheat genome. This induced translocation event normally comprises of a piece of alien chromosome with many neighboring donor genes, including the desired resistance gene. Linkage drag is also in this case a negative side effect from the method used. If negative traits are linked to the gene of interest, it is not easy to remove them by meiotic recombination, which is suppressed at such a location. The same holds true for the introgressed piece of alien chromosome after suppression of homoeologous chromosome pairing during the absence of the *Ph* gene.

The cisgenic approach could also in this case be very helpful. It is a one step approach instead of all complicated steps described above. Precondition is the availability of the gene of interest and transformation ability of the variety which has to be improved. In the case of complex allopolyploid crops the advantage of cisgenesis is highly evident. It will improve breeding power of this type of crops considerably.

25.4.3 Cisgenesis and Its Potential Use in Further Breeding

After cloning of the alleles coding for the desired traits, cisgenesis is an alternative for cross breeding as well as introgression and induced translocation breeding. It is simplifying, as indicated in Fig. 25.2, normal introduction, introgression or translocation of such traits. It is specifically of interest if more than one trait, coming from different breeding parents or species, need to be combined at the same moment. As first step an existing transformable top variety, which has to be improved, is needed. If several cisgenes are introduced, using a gene cassette with these cisgenes, then during subsequent cross breeding these cisgenes are genetically tightly linked together. This is another major advantage. In this way, quantitative traits with a polygenic basis can obtain more attention and be more extensively handled in cross breeding.

Table 25.2 Breeding characteristics of some major crops

Crop species	Mode of reproduction	Type of cultivar	Homogeneity	Zygosity
Apple	Vegetative	Cultivars	Homogeneous	Heterozygous
Asparagus	Allogamous (dioecious)	open pollination	Heterogeneous	Heterozygous
		F1 Hybrids	Homogeneous	Heterozygous
Barley	Autogamous	Landrace	Heterogeneous	Homozygous
Wheat		Modern Cultivars	Homogeneous	Homozygous
Cucumber	Allogamous monoecious	F1 hybrids	Homogeneous	Heterozygous
Maize	Allogamous (monoecious)	Landrace and open pollination	Heterogeneous	Heterozygous
		F1 hybrids	Homogeneous	Heterozygous
		Inbred lines	Homogeneous	Homozygous
Potato	Vegetative	Cultivars	Homogeneous	Heterozygous
	Sexual	True potato		
		Seed	Heterogeneous	Heterozygous
Rye	Allogamous	Open pollination	Heterogeneous	Heterozygous
Tomato	Autogamous	Old/modern	Homogeneous	Homozygous
Rice		Cultivars	Homogeneous	Heterozygous
		F1 hybrids		

25.5 Hybrid Varieties in Allogamous Crops

Many allogamous crops are more and more pushed into the direction of hybrid varieties. The classical way of breeding is based upon open pollination. In the allogamous crops we will restrict to breeding for hybrid seed varieties. Table 25.2 is representing 1. autogamous crops like barley with landraces and modern varieties and tomato with old varieties and F1 hybrids, 2 allogamous crops, a. vegetatively propagated crops like apple and potato with cultivars based on one genotype, and b. seed propagated crops with landraces, open pollinated varieties, F1-hybrids and inbred lines. The allogamous crops can be subdivided in monooecious like cucumber and maize or dioecious like asparagus. For hybrid seed production male sterility, based on GMS (genetic male sterility) or CMS (cytoplasmic male sterility), and/or gametophytic or sporophytic self-incompatibility are important traits.

25.5.1 Male Sterility in Traditional Breeding

In nature many types of male sterility are found. For breeding the most useful form is based on nucleocytoplasmic male sterility (CMS). In, for example, *Brassica* vegetables like Chinese cabbage, which are grown for their vegetative parts, the hybrid plant can still be male sterile without influencing yield negatively. In other crops like maize and rice, where seeds are the end product, restoration of male fertility is needed. In such cases, restorer genes (*Rf*) are very important. In practise, these *Rf* genes are complicating variety development considerably. In recent years a lot of knowledge has been obtained about the molecular genetic base of CMS as well as *Rf* genes (Wise and Pring 2002). This brings the possibility of manipulating CMS and its restoration by the GM approach very near, specifically when cisgenes could be involved. In several species like maize (T-*urf13)*, bean (*pvs-orf239*), *Brassica* (*orf138*) and *Petunia* (*pcf*) the CMS-genes involved have been identified (Pelletier and Budar 2007; Bentolila 2002). These genes are in all cases resulting from complex DNA rearrangements of mitochondrial DNA. It is remarkable that here single mitochondrial genes are involved in male sterility. These natural MS-genes are available for GM-approaches and genetic modification of mt-DNA is the best option, however, mt-DNA based transformation has still to be developed.

In seed propagated crops, like maize and rice, hybrid varieties are dependent on the presence of a restoration system. In breeding of hybrid varieties of maize, *Rf* genes have been used already for many years. However, as shown in Fig. 25.3 it is always complicating the breeding system considerably**.** In addition to the mother line, production of the hybrid variety by using the CMS mother line in cross combination with the fertile *Rf* free maintainer line, which contains normal cytoplasm, it is important to have a *Rf* containing father line for making a

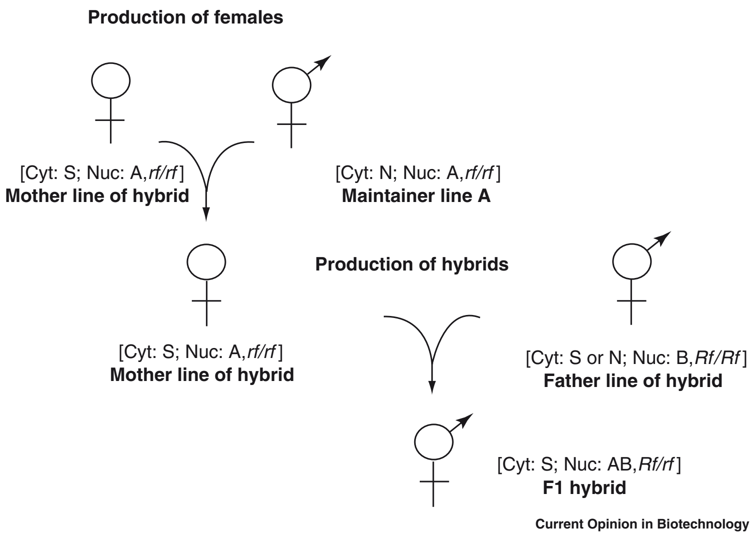


Fig. 25.3 The production of male fertile F1 hybrid seed by the use of cytoplasmic male sterility in cms Mother lines, in combination with male fertile Maintainer lines with normal cytoplasm and male fertile Father lines with normal or cms cytoplasm which are homozygous for the restorer gene (*Rf*) (Pelletier. and Budar 2007)

fertile hybrid. Crossing of this father line with the CMS mother line delivers male fertile hybrid seed with CMS cytoplasm but heterozygous for the restorer gene (*Rfrf*). In recent years, *Rf* genes of Petunia (*Rf-PPR592*; Bentolila et al. 2002), rice (*Rf1A* and *Rf1B*; Wang et al. 2006) and maize (*Rf2-R213*; Liu et al. 2001) have been isolated and can be used in cisgenic GM-approaches.

25.5.2 Transgenesis for Introduction of Male Sterility

Male sterility has already been manipulated for a long time by using transgenes. Most important example is the BAR-STAR system (Reynaarts et al. 1993), successfully applied in *Brassica napus,* and used in combination with herbicide resistance. With the *Barnase* gene (an RNase gene from *Aspergillus oryzae*) GMS can be introduced by using a tapetum specific promoter from tobacco. The *Barstar* gene, (an RNase inhibitor gene from *Bacillus amyloliquefaciens*) also driven by the tapetum promoter, is inhibiting the *Barnase* gene and restoring male fertility. The use of both transgenes is enabling hybrid seed production system in cultivated plants, providing *Brassica napus* a high increase in yield. However, the disadvantage is, that this system is fully based on transgenes.

25.5.3 Cisgenesis and Introduction of Male Sterility

Classical hybrid seed production has been applied in more and more crops. As shown above, the use of CMS in combination with restoration of fertility takes a lot of energy and breeding effort.Manipulation of both traits by using cisgenes is possible, as has been shown above by the molecular isolation of natural mt-based ORF-genes causing male sterility, and of natural nuclear *Rf*-genes restoring male sterility. The GM-approach would bring much more flexibility to the classical breeder by adding both traits directly to existing, good combining, lines. The GM-approach with *Rf*- cisgenes is already possible now. The introduction of *ORF*-cisgenes causing CMS into mt-DNA is not possible now, however, introducing them into chloroplast-DNA is today already a realistic option. The first example of engineering male sterility via transformation of the chloroplast genome is expression of a transgene coding for beta-ketoliase (Ruiz and Daniell 2005). A remarkable observation was that reversibility of the CMS phenotype could be obtained under continuous illumination of the transgenic tobacco plant. In the cases where engineering of mt-DNA and/or chloroplast-DNA transformation is involved, a GM-plant is obtained, however, in case the genes used belong to the existing breeders gene pool it should be exempted. In case of the availability of cisgenic *Rf* genes the same breeding scheme could be followed as shown in Fig. 25.3 however, the *Rf* containing father line genes could be obtained more easily.

25.5.4 Self-Incompatibility and a Possible Role of Cisgenesis

Self-incompatibility is promoting outcrossing. It is one of the most important ways in plants to prevent self-fertilization and inbreeding. For breeding of cross-fertilizing crops, self-incompatibility can be a serious hindrance and in plant breeding there are mainly two systems of self-incompatibility, i.e. gametophytic (GI) and sporophytic (SI) self-incompatibility. The GI system can be based on one locus (cabbage, red clover) with many alleles or on two loci (grasses). The gametophytic one-locus self-incompatibility is molecularly well investigated. It results from the interaction between the haploid genotype of the pollen grain and the diploid genotype of the style. Inhibition is found in the style. The *S*-alleles in the style work independently. It is known that polyploidization restores self-compatibility.

Plants belonging to the *Compositae* family do possess the SI system. This is usually based on one gene with multiple alleles. The interaction is dependent on the diploid maternal genotype of the pollinator and the diploid genotype of the style. Inhibition is found on the stigma. Reaction of the pollen is determined by the dominance relationship of the sporophyte and not by the gametophyte. For hybrid breeding the SI system is frequently used in vegetable crops of *Brassica oleracea.*

Molecular research on GI of, for example, *Petunia inflata* brought new possibilities for influencing self-incompatibility (McClure 2006). The incompatibility system is complicated but is S-RNase based. *Petunia* transformed with the *PiSLF2* cisgene showed in a heteroallelic background, as earlier observed after polyploidization, breakdown of GI but not in a homoallelic background. It is clear from recent

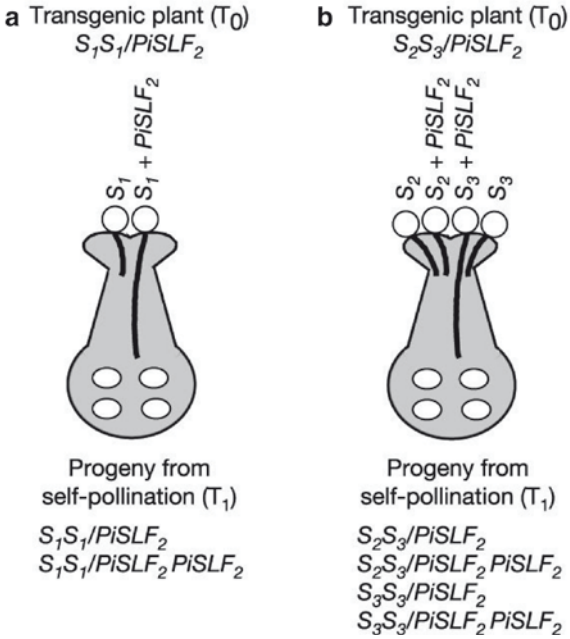


Fig. 25.4 Schematic representation of transformation experiments to ascertain the function of *PiSLF*. (**a**) Self-incompatibility behaviour of an *S1S1* transgenic plant carrying a single copy of the *PiSLF2* transgene. (**b**) Self-incompatibility behaviour of an *S2S3* transgenic plant carrying a single copy of the *PiSLF2* transgene. The genotypes of pollen produced, the predicted *S*-genotypes of the progeny resulting from self-pollination, and inheritance of the transgene are indicated (Sijacic et al. 2004)

literature that unraveling of the whole GI system is complicated but at the other hand the results with the cloned *S*-alleles in GM-plants of *Petunia inflata* (Sijacic et al. 2004), as seen in Fig. 25.4, showed that these cisgenes can be used for inhibiting self-incompatibility. This possibility is based on the earlier observation that two different *S*-alleles present in one pollen grain prevent the incompatibility reaction in the style. It means that for hybrid varieties, inbreeding is more feasible with additional *S*-cisgenes. If transformation is not an obstacle, these cisgenes could be used for developing inbred lines and/or seed producing hybrid varieties.

25.6 Vegetatively Propagated Crops

25.6.1 Traditional Breeding

There are many important vegetatively propagated crops like potato, cassava and other root and tuber crops, fruit crops like apple, pear, plumb, peach, banana, ornamentals like Chrysanthemum, Alstroemeria and rose, and bulb crops like tulip, lily,

onion (shallot). All these crops have in common that varieties are consisting of one genotype which is vegetatively propagated by roots, tubers, bulbs, shoots, rhizomes or grafts. These crops are annual or perennial. They have also in common that these crops are genetically highly heterozygous and that crosses are recombining all combined traits immediately so that the parental genotype is lost and can never come back in the same combination via backcross procedures. In all these crops it is very well known that variety breeding demands increasingly numbers of seedlings in order to be able to find seedlings combining more desired traits. Nowadays, in potato over 1,00,000 seedlings are needed for one new variety. Genetic variation has also to come more frequently from wild species so that pre-breeding for breeding parents is a very important first step in the breeding process using wild plants. Success of pre-breeding is in these cases also highly dependent on linkage drag problems with presence of undesired alien traits.

It is important to realize that improvement of existing varieties in traditional breeding is only possible by spontaneous mutations or by mutation induction. This way of improvement is frequently used in fruit crops like apple and in ornamentals like Chrysanthemum, Alstroemeria and rose and bulb species like tulip, crocus and narcissus, where changes in appearance are directly useful. Loss of function of specific genes can also be obtained by the RNAi approach, This is very powerful as has been shown, for example, for amylose-free potato starch (Heilersig et al. 2006). However, the outcome is a GM-plant which is based on a transgene, which could nowadays be indicated as an intragene if all functional parts are coming from natural gene sequences of the crop plant itself or from crossable species. In case of an RNAi intragene the deregulation could also be made less costly and less time consuming because all functional parts of the intragene are belonging to genes from the plant itself or from crossable species.

25.6.2 Extension of Genetic Variation

In agricultural practice and post harvest technology gain of functions, such as those influencing quality or resistances to biotic stress factors are very important. These traits can be found e.g. in wild material. The most important problems that have to be solved in this way are susceptibility to many diseases. In apple, resistance to apple scab (*Venturia inaequalis*) is at this moment a hot item and in potato resistance to late blight, nematodes and wart disease. Resistances to these diseases have to come from wild species and in practice several genes resisting different isolates of the same pathogen have to be combined, which is called gene stacking. These two examples of potato and apple will be discussed below in more detail.

25.6.3 Potato–Phytophthora Interaction

In a case like *Phytophthora infestans* it is very well known from the past that resistance can be overcome by the pathogen relatively easily (Haverkort et al. 2008).

Until the fifties *Solanum demissum* was used as main source of resistance to *Phytophthora* without sustainability in success. All 11 *R*-genes from *S. demissum* were broken and attention was shifted to horizontal resistance which is more based on quantitative traits. In potato this approach appeared also to be without sufficient amounts of success. Only a few varieties have been obtained with quantitative resistance. The present solution is searching for resistance in many different crossable species. These resistances do segregate well in crosses with susceptible plants and are effective against different complex isolates. In this way, *R*-genes are selected which in combination with other major resistance genes could be the basis for more durable resistance (Jacobsen and Vossen 2009). Sources of resistance in different species are found and needed, which brings more often introgression problems because of a lower degree of crossability and linkage drag problems because of decreased cross-over events in the homoeologous parts of the chromosomes of backcross plants during meiosis. Introgression breeding in vegetatively propagated crops brings the same problems of linkage drag as shown above for self-fertilizing crops. Fig. 25.5 shows how many years are needed for a new variety in traditional breeding when (double) bridge interspecific crosses are required to transfer major resistance genes into normal breeding material. In case of bridge crosses with far related species it is expected that during backcrossing meiotic cross-over events occur less frequently in the introgressed areas. This is generally increasing the size of introgressed alien chromosome parts and the chance of negative side effects by other linked alien genes. In case of potato, it has taken over 50 years before varieties with the first *R*-genes of *S. bulbocastanum* against *Phytophthora* could be released as recently has been done with the new dutch varieties Toluca and Bionica (personal comm.). This highly resistant variety probably contains only *Rpi-blb2* (unpublished observations) which still provides in the Netherlands total resistance however, in trap fields virulent isolates of late blight have already been isolated (Kessel, pers. comm). Therefore, only one *R*-gene in a variety is dangerous because of potential breakage problems. The philosophy has to be to stack several major *R*-genes from different sources at the same time in one genotype (Jacobsen and Vossen 2008). In classical breeding this leads to accumulation of linkage drag problems and, therefore,

Double-bridge crosses in potato for resistance to *Phytophthora infestans*

S. acaule 4x × *S. bulbocastanum* 2x *(R genes)*
↓
AB hybrid 3x
↓ colchicine doubling
AB hybrid 6x × *S. phureja* 2x
↓
ABP hybrid 4x × *S. tuberosum* 2x
↓
ABPT material 4x *R*-gene + linkage-drag

Fig. 25.5 Double-bridge crosses in potato introgression breeding for resistance to *Phytophthora* infestans. This multiple step approach with far related species is difficult and always accompanied with a lot of linkage drag around the donor resistance gene. Stacking of *R*-genes from different sources, without linkage drag problems, is complex

decreasing the chance of providing attractive varieties in a reasonable time. The tetraploid and heterozygous nature of potato further complicates the combination of desired traits in a new variety.

Alternatives like stacking durable *R*-genes via cisgenesis are in development in order to solve this problem. At the moment 15 *R*-genes from several species like *S. demissum, S. bulbocastanum, S. stoloniferum, S. papita* and *S. venturii* have been isolated and are tested with different isolates for their spectrum of resistance. The use of *Avr*-genes is also speeding up the detection of *R*-genes in other species with the same or nearly the same specificity. It is getting more and more clear that *R*-genes can be subdivided into different groups with various resistance mechanisms (Vleeshouwers et al. 2008). These isolated *R*-genes can be stacked more easily by cisgenesis, and the availability of the corresponding *Avr*-genes enable testing the simultaneous biological expression of these *R*-genes in one plant. The possibility of using *R*-genes in GM-plants is also promoting the development of new resistance strategies as has been applied in the past with *Bt* based insect resistance genes by using refuges (Babu et al. 2003). In potato three resistance strategies have been described which could be used in practice (Jacobsen and Vossen 2008). It is one by one using single *R*-genes, stacking *R*-genes in one variety or using different *R*-genes in different clones from the same variety in mixed varieties.

25.6.3.1 Apple–Venturia Inaequalis Interaction

In 1946 crosses were made for introduction of resistance to apple scab (*Venturia inaequalis*) into commercial apple varieties, using as source of resistance the crab apple *Malus floribunda* 821 (Hough et al. 1953). The progeny of the cross between *M. floribunda* 821 and susceptible cultivars segregated in a Mendelian fashion for resistance in a 1:1 ratio. The gene putatively underlying this resistance was named *Vf*-gene. However, the fruits of the resistant parent *M. floribunda* 821 were very small, approximately 1 cm. The apples of the progeny were also small, and did not have the fruit quality that was required for commercial cultivars. This was caused by linkage drag: not only the desired resistance gene was inherited to part of the progeny, but also many unwanted alleles leading to poor fruit quality and other unwanted traits. In order to get rid of the unwanted alleles, subsequent crosses had to be carried out between resistant progeny and susceptible high quality cultivars. As indicated in Fig. 25.6 about five generations were required to remove enough unwanted alleles from *M. floribunda*, yet keeping the desired *Vf*-gene for scab resistance. Approximately 50 years after the first cross, *Vf*-cultivars with a reasonable fruit quality were introduced onto the market (Anonymous 1999). Therefore, it has taken half a century to introduce the *Vf*-gene and remove the linkage drag to an acceptable degree.

In the mean time, *Venturia inaequalis* strains have been detected that are able to infect *Vf*-cultivars (Parisi et al. 1993). Especially in North-western Europe these strains are present and have spread (Parisi et al. 2006). As a result, several orchards that consist of *Vf*-cultivars have to be sprayed like orchards with susceptible

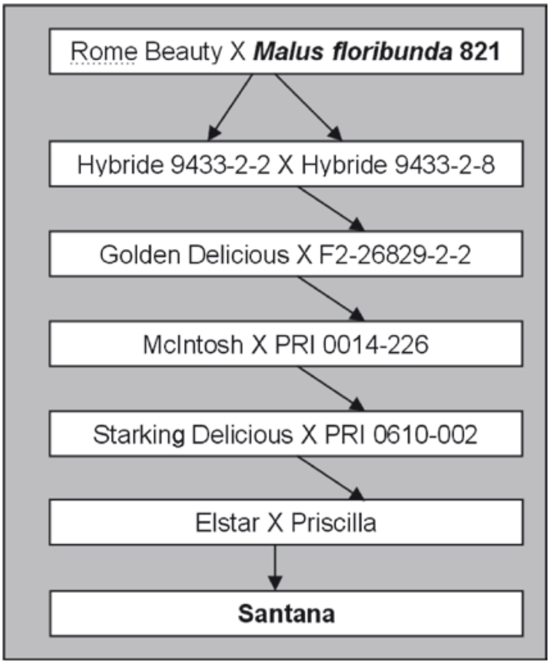


Fig. 25.6 Breeding scheme between apple and Malus *floribunda* 821 for introgression of the Vf resistance gene against apple scab (*Venturia inaequalis*) providing after 50 years the resistant cultivar Santana

cultivars (Trapman 2006). Fifty years of breeding is fading away in 10 years. Obviously, more individual resistance genes need to be accumulated for obtaining durable resistance.

Fortunately, many loci that confer resistance to apple scab have been discovered in *Malus*, both major genes and QTLs (Calenge et al. 2004; Schmidt and Van de Weg 2005; Gessler et al. 2006; Gardiner et al. 2006). Therefore, sufficient genes for resistance are present in the germplasm of apple. Introgression of one resistance gene took approximately 50 years. Introgression of four or more genes for durable resistance will require more time, when the breeding is performed in the classical way. Would this imply that we are left with an additional 50 years of intensive fungicide applications to scab in apple, before the durably resistant cultivars are introduced? We regard it as our challenge to shorten this period significantly, by introducing the resistance genes and preventing the linkage drag.

One way to speed up the breeding process is marker-assisted breeding. This method can be of tremendous use and we advocate this approach, but still this will be time-consuming, because linkage drag has to be removed through crosses and meiotic recombination. Accumulation of resistance genes from four sources of resistance and sufficient removal of unwanted alleles from the same sources, will probably require at least another five cross generations of apple. As long as the juvenile

period and additional evaluation time in apple is about eight years, this would require a minimum of 40 years of breeding.

An alternative route is introduction of the resistance genes into susceptible elite cultivars without simultaneous introgression of unwanted alleles, so prevention of linkage drag, rather then removal of linkage drag. Through this process of cisgenesis, durable resistance provided by several resistance genes is added to high quality cultivars in one step, preserving the proven fruit quality and other desired traits of these cultivars.

Currently, apple is being sequenced, and other fruit crops probably will soon follow. This provides unprecedented opportunities for identification of genes. In addition, numerous loci have been mapped genetically in diverse germplasms, including fruit crops (Kole 2006). The information on genetic positions on the linkage groups, together with the whole genome sequences, and knowledge of genes from model plant species, offer us great opportunities to isolate alleles for desired traits at an increasing efficiency. We expect that, in the coming ten years, a vast number of major alleles for desired traits will be isolated in many crops, including fruit. So, the treasury of isolated alleles for cisgenesis will be filled at an increasing rate.

The already mentioned *Vf*-gene was isolated by means of map-based cloning and subsequently functionally analyzed (Belfanti et al. 2004). A closer analysis revealed that a tandem repeat of two gene copies provides the *Vf*-resistance (Malnoy et al. 2007). Therefore, the spelling '*Vf*-gene' should be updated to the plural form '*Vf*-genes'. In the mean time several other resistance genes to apple scab are being isolated. As soon as the apple genome sequence becomes available to the scientific community, many more genes and their alleles will also be isolated and characterized, and will enrich the wealth of available alleles for cisgenesis.

After discovery of the *Vf*-genes, several research groups in Europe and the USA proceeded in inserting these genes with strong, constitutive promoters in susceptible cultivars, resulting in resistance. However, at nearly the same time, the natural cisgenes were inserted with their own promoters. Apparently, the concept of cisgenesis was a logical step following the isolation of the *Vf*-genes.

Additional resistance genes will also be inserted into apple by means of cisgenesis within a few years, using a combination of *Vf*-genes and other resistance genes. These stacked functional resistance genes will provide more durable resistance in elite cultivars. We regard cisgenesis as a way to apply the increasing knowledge about alleles to plant breeding, to the benefit of growers, consumers and the environment.

An extra advantage of cisgenesis in comparison with cross breeding is that susceptible cultivars can be used that already have a proven high fruit quality and safe use. Apple is self-incompatible. Crossing with apple germplasm scrambles the genetic composition of good cultivars, and restoring such a cultivar through crossings is virtually impossible. However, cisgenesis preserves the genetic assembly of the high quality cultivar, and adds some well-defined apple alleles. Subsequently, the enriched cultivar can be propagated vegetatively by means of grafting, which is a common practise in apple propagation.

25.6.4 Cisgenic Approach in Vegetatively Crops

The stacking and linkage drag problems are, in addition to allopolyploids, most prominent in vegetatively propagated crops, specifically when they are polyploid or perennial as has been described in potato and apple, respectively. Short run success can be obtained in this type of crops when cisgenes are available and can be transferred directly into the variety of interest. During stacking of genes it is important to know whether all the genes of the cassette after insertion are expressed and biologically active. For strong resistance genes which have not been broken at the moment it is important to know whether HR reactions with the cross reacting *Avr*-genes is occurring or not and whether this is a reliable indication for functional gene expression. The cisgenic approach is, as earlier indicated, free of the multiple introgression problems when the traditional breeding approach has to be applied so that domestication of a number of inserted alien genes can be realized in one step. It is also important to realize that stacked cisgenes in existing varieties can easily be used in traditional cross breeding approaches. We believe that the cisgenic approach is simplifying traditional plant breeding, especially when multiple linkage drag problems have to be overcome. If this approach will be exempted from the GM-regulations worldwide, it will help plant breeding in general.

25.7 Concluding Remarks

This chapter describes cisgenesis as an efficient alternative for several multiple step (pre-breeding) approaches in traditional plant breeding by highlighting with examples for important traits that. It is introducing and domesticating agriculturally important genes in improved existing varieties in one step and providing new crossing parents for variety breeding. This approach is, therefore, highly recommended for breeders in developing countries and for SMEs. The process of producing cisgenic GM-plants results per definition into a GM-plant for the same reason as indicated in the GM-Directive 2001/18/EC (Anonymus 2001) for induced mutations and protoplast fusion between crossable species. However, in parallel with the induced mutation and protoplast fusion approaches, the cisgene source belongs per definition to the well known and already for a long time used breeders gene pool and should, because of this, also be exempted from the existing GM-regulations worldwide.

References

Anonymous (1999)Variety list of fruit crops (in Dutch), Dutch Fruit Growers Organization, P.O. Box 906071, 2585 HK The Hague, The Netherlands, ISN 0169–6750

Anonymous (2001) Directive 2001/18EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC. Off. J Eur Comm 106:1–38 (http://europa.eu.int/eur-lex/pri/en/oj/dat/2001/1_10620010417en00010038.pdf)

Babu RM, Sajeena A, Seetharaman K, Reddy MS (2003) Advances in genetically engineered (transgenic) plants in pest management–an overview. Crop Prot 22:1071–1086
Belfanti E, Silfverberg-Dilworth E, Tartarini S, Patocchi A, Barbieri M, Zhu J, Vinatzer BA, Gianfranceschi L, Gessler C, Sansavini S (2004) The HcrVf2 gene from a wild apple confers scab resistance to a transgenic cultivated variety. ProcNatl Acad Sci U S A 101: 886–890
Bentolila S, Alfonso A, Hanson MR (2002) A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. Proc Natl Acad Sci 99:10887–10892
Buckler E, Thornsberry JM, Kresovich S (2001) Molecular diversity, structure and domestication of grasses. Genet Res Cambridge 77:213–218
Calenge F, Faure A, Goerre M, Gebhardt C, Van de Weg WE, Parisi L, Durel C-E (2004) Quantitative Trait Loci (QTL) analysis reveals both broad-spectrum and isolate-specific QTL for scab resistance in an apple progeny challenged with eight isolates of Venturia inaequalis. Phytopathology 94:370–379
Friebe B, Jiang J, Raupp WP, McIntosh RA, Gill BS (1996) Characterization of wheat-alien translocations conferring resistance to diseases and pests: current status. Euphytica 91:59–87
Gardiner SE, Bus VGM, Rusholme RL, Chagné D, Rikkerink EHA (2006) Apple. In: Kole C (ed) Genome mapping and molecular breeding. Vol 4, Fruit and nuts. Springer, Berlin, pp 2–62
Gessler C, Patocchi A, Sansavini S, Tartarini S, Gianfranceschi L (2006) Venturia inaequalis resistance in apple. Crit Rev Plant Sci 25:473–503
Haverkort AJ, Boonekamp PM, Hutten R, Jacobsen E, Lotz LAP, Kessel GJT, Visser RGF, Vossen Van der EAG (2008) Societal costs of late blight in potato and prospects of durable resistance through cisgenic modification. Potato Res 51:47–57
Heilersig HJB, Loonen AS, Bergervoet M, Wolters AMA, Visser RGF (2006) Post-transcriptional gene silencing of GBSSI in potato: effects of size and sequence of the inverted repeat. Plant Mol Biol 60:647–662
Hough LF, Shay JR, Dayton DF (1953) Apple scab resistance from Malus floribunda Sieb. Proc Am Soc Hort Sci 44:260–272
Jacobsen E (2007) Cisgenesis: next step in advanced traditional breeding. In: Haverkort AJ, Anisimov BV (eds) Potato production and innovative technologies.Wageningen Academic, ISBN:978–90–8686–042–5, pp 348–352
Jacobsen E, Schouten HJ (2007) Cisgenesis strongly improves introgression breeding and induced translocation breeding of plants. Trends Biotechnol 25: 219–223
Jacobsen E, Vossen, van der EAG (2009) Plant disease resistance. Cisgenesis: a new tool for traditional resistance breeding. In: Encyclopedia of microbiology, 3rd edn. Mesolio Schaechter Editor) pp 597–612
Jansen JPA (1997) Aphid resistance in composites. Patent NL1003261
Jarvis DJ, Hodgkin T (1999) Wild relatives and crop cultivars: detecting natural introgression and farmer selection of new genetic combinations in agroecosystems. Mol Ecol 8:159–173
Kole, C. (2006) Genome mapping and molecular breeding in plants. Vol 4, Fruits and nuts. Springer, Berlin
Lim KY, Leitch IJ, Leitch AR (1998) Genomic characterisation and the detection of raspberry chromatin in polyploid Rubus. Theor Appl Genet 97:1027–1033
Liu F, Cui X, Horner HT, Weiner H, Schnable PS (2001) Mitochondrial aldehyde dehydrogenase activity is required for male fertility in maize. Plant Cell 13:1063–1078
Malnoy M, Xu M, Borejsza-Wysocka EE, Korban SS, Aldwinckle HS (2007) The role of Vfa1, Vfa2 and Vfa4 genes in resistance to Venturia inaequalis in transgenic ‘McIntosh’ apple plants. ISHS Acta Hort 738:323–326
McClure B (2006) New views of S-RNase-based self-incompatibility. Curr Opin Plant Biotechnol 9:639–646
Parisi L, Lespinasse Y, Guillaumes J, Kruger J (1993) A new race of Venturia inaequalis virulent to apples with resistance due to the Vf gene. Phytopathology 83:533–537
Parisi L, Laurens F, Didelot F, Evans K, Fischer C, Fouillet V, Gennari F, Kemp H, Lateur M, Patocchi A, Schouten HJ, Tsipouridis C (2006) Geographical distribution of Venturia inaequalis strains virulent to the Vf gene. Europe. Bulletin-OILB/SROP 29:49–52

Pelletier G, Budar F (2007) The molecular biology of cytoplasmically inherited male sterility and prospects for its engineering. Curr Opin Biotechnol 18:121–125

Reynaarts A, Wiele van der H, Sutter de G, Janssens J (1993) Engineered genes for fertility control and their application in hybrid seed production. Scientia Horticulturae 55:125–139

Rommens CM (2004) All-native DNA transformation: a new approach to plant genetic engineering. Trends Plant 9:457–464

Rommens CM, Ye J, Richael C, Swords K (2006) Improving potato storage and processing characteristics through all-native DNA transformation. J Agric Food Chem 54:9882–9887

Rommens CM, Haring MA, Swords K, Davies HV, Belknap WR (2007) The intragenic approach as a new extension to traditional plant breeding. Trends Plant Sci 12:397–403

Ruiz ON, Daniell H (2005) Engineering cytoplasmic male sterility via the chloroplast genome by expression of beta-kethiolase. Plant Physiol 138:1232–1246

Schaart J, Krens FA, Pelgrom KTB, Mendes O, Rouwendaal GJA (2004) Effective production of marker-free transgenic strawberry plants using inducible site-specific recombination and a bifunctional selectable marker gene. Plant Biotechnol J 2:233–240

Schmidt H, Van de Weg WE (2005) Breeding. In: Tromp J., Webster AD, Wertheim SJ (eds) Fundamentals of temperate zone tree fruit production. Backhuys, Leiden, the Netherlands, pp 136–155

Schmidt R (2002) Plant genome evolution: lessons from comparative genomics at the DNA level. Plant Mol Biol 48:21–37

Schouten HJ, Krens FA, Jacobsen E (2006a). Do cisgenic plants warrant less stringent oversight? Nat Biotechnol 24:753

Schouten HJ, Krens FA, Jacobsen E (2006b). Cisgenic plants are similar to traditionally bred plants. EMBO Rep 7:750–753

Sijacic P, Wang X, Skirpan AL, Wang Y, Dowd PE, McCubbin AG, Huang S, Kao TH (2004) Identification of the pollen determinant of S-RNase-mediated self-incompatibility. Nature 429:302–305

Trapman, M. (2006) Resistance management in Vf resistant organic apple orchards. Bulletin OILB/SROP 29:253–257

Vetten de N, Wolters AMA, Raemakers K, Meer van der I, Stege ter R, Heeres E, Heeres, Visser RGF (2003) A transformation method for obtaining marker-free plants of a cross-pollinating and vegetatively propagated crop. Nat Biotechnol 21:439–442

Vleeshouwers GAA, Rietman H, Krenek P, Champouret N, Young C, Oh SK, Wang M, Bouwmeester K, Vosman B, Visser RGF, Jacobsen E, Govers F, Kamoun S. Vossen, van der EAG (2008) Effector genomics accelerates discovery and functional profiling of potato disease resistance and Phytophthora infestans avirulence genes. PLoS ONE e2875

Wang Z, Zou Y, Li X, Zhang Q, Chen L, Wu H, Su D, Chen Y, Guo J, Luo D, Long Y, Liu YG (2006) Cytoplasmic male sterility of rice with Boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. Plant Cell 18:676–687

William HM, Trethowan R, Crosby-Galvan EM (2007) Wheat breeding assisted by markers: Cimmyts experience. Euphytica 157:307–319

Wise RP, Pring DR (2002) Nuclear-mediated mitochondrial gene regulation and male fertility in higher plants: light at the end of the tunnel? Proc Natl Acad Sci 99:10240–10242

Yu HX, Tang SZ, Liu QQ, Wang L, Zhao ZP, Xu L, Huang BL, Gong ZY, Tang SZ, Gu MH (2006) Breeding of selectable marker-free transgenic rice lines containing AP1 gene with enhanced disease resistance. Agric Sci Chin 5:805–811

Chapter 26
Gene Stacking

E. Douglas and C. Halpin

Abstract A small but increasing proportion of genetically modified crops harbour two or more novel traits due to 'stacked' transgenes. A variety of methods can be used to achieve stacking, albeit with limitations. Transgene stacking can potentially widen the scope of current plant genetic manipulation to allow whole new biochemical pathways to be introduced into plants, or to overcome a range of different factors that limit crop yield. Developing and improving methods for multi-gene stacking in plants is an expanding and exciting field of current research.

26.1 Introduction

The manipulation of plant genomes via genetic modification offers the potential to improve crop plants in many beneficial ways. Plants can be engineered to have increased nutritional benefits, resistance to herbicides and insecticides and enhanced tolerance to both biotic and abiotic stresses, as well as being able to produce pharmaceutical products such as vaccines. Commercial transgenic crop production continues to gain ground with 114 million hectares grown globally in 2007 (James 2007; Fig. 26.1a). Although a single engineered trait still dominates (63% of biotech crops planted were herbicide tolerant), the fastest growing category between 2006 and 2007 was crops that combined two or more 'stacked' traits. Within this growing season, the area planted with crops with stacked traits rose 66% to cover 22 million hectares (i.e. 19% of global transgenic crop production) and over a third of all GM crops grown in the United States possessed two to three stacked traits (James 2007). The adoption of stacked trait crops appears to be entering a rapid growth phase (Fig. 26.1b) likely reflecting the potential of these

E. Douglas
Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK

C. Halpin (✉)
Division of Plant Sciences, College of Life Sciences, University of Dundee at SCRI, Invergowrie, Dundee, DD2 5DA, UK
e-mail: chalpin@dundee.ac.uk

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_26,

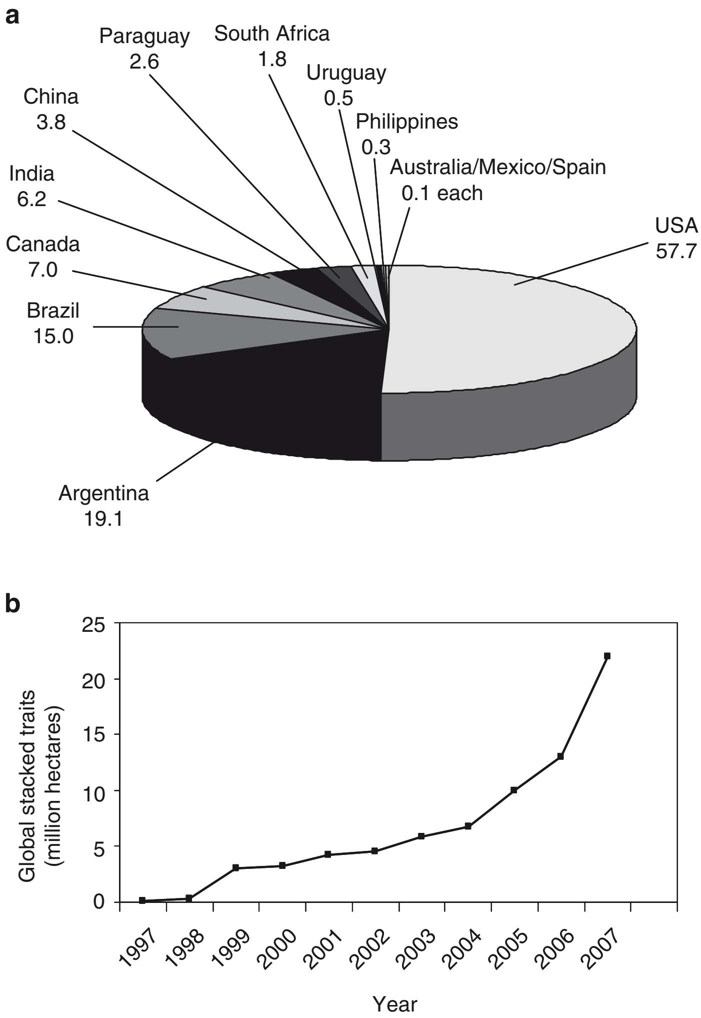


Fig. 26.1 (**a**) The top 13 countries growing most GM crops in 2007 (numbers in million hectares). Drawn from data in James 2007. (**b**) Global area of stacked trait transgenic crops by year over the past decade. Data from James 1997–2007

crops to overcome multiple factors limiting crop productivity (Halpin 2005). Already, varieties of GM maize and cotton are available with three stacked traits, and Bayer CropScience and Monsanto recently announced their intention to produce the first-ever eight-gene stacked commercial crop resistant to several insects and herbicides. Despite these advances, the underpinning techniques used to achieve multiple trait/gene stacking in commercial transgenic crops, although effective, are still relatively basic. This chapter discusses those techniques along with newer research that could potentially provide more efficient technologies for multiple gene stacking in the future.

26.2 Combining Individual Transgenes

26.2.1 Cross-Breeding and Re-Transformation

The introduction of multiple novel genes into plants (stacking or 'pyramiding') can be achieved by straightforward iterative strategies. Plants containing several transgenes can be produced by crossing parents with different transgenes until all the required genes are present in the progeny (Fig. 26.2a). This has been the method used to produce most of the three-gene stacked GM crops currently available such as maize that is corn borer and rootworm resistant and herbicide tolerant. At a research level, cross-breeding has been used to introduce novel proteins or new biochemical pathways into plants. An early example of the power of this strategy was the production of secretory IgA antibodies in plants by cross-breeding of tobacco to combine, in one plant, four genes encoding different immunoglobulin polypeptides (Ma et al. 1995). Similarly, Arabidopsis plants tolerant of the environmental pollutant methylmercury were produced by crossing plants expressing two different genes on a bacterial organic mercury detoxification pathway (Bizily et al. 2000).

An alternative iterative strategy is to sequentially transform and re-transform plants with different individual transgenes by several rounds of transformation (François et al. 2002a; Halpin and Ryan 2004) (Fig. 26.2b). This strategy can be particularly useful in crops that are not easy to propagate by sexual crossing, such as woody plants and trees. In forsythia, flower colour has been modified by sequential transformation with the genes for dihydroflavonol 4-reductase from Antirrhinum majus and anthocyanidin synthase from Matthiola incana This induced anthocyanin synthesis in the double transformants which displayed a novel bronze-orange petal colour (Rosati et al. 2003). Re-transformation has also been used in poplar to combine antisense transgenes for two lignin biosynthesis genes in an attempt to reduce the environmental impact of papermaking (Lapierre et al. 1999).

Both of these iterative methods allow the effect of each additional gene to be assessed in a step-wise fashion, which may offer advantages in some circumstances. However, whilst these methods reliably result in the production of plants expressing multiple transgenes, problems can still arise. Principal among these is the fact that the introduced transgenes will be integrated randomly in different genomic positions which can result in (1) lack of coordination between the expression levels of different transgenes, and (2) subsequent independent segregation of the transgenes in later generations. Both techniques are fairly labour intensive, taking several generations to complete (Halpin et al. 2001; François et al. 2002a). An additional issue, particularly with the re-transformation strategy, is that different selectable marker genes may be needed for each sequential round of transformation, and the availability of such genes can be limiting. The accumulation of different selectable marker genes as more transgenes are introduced could also, potentially, hinder regulatory approval and public acceptance of the final product. The development of systems enabling transformation without the need for selectable markers, or systems for removal of selectable marker genes, may help overcome these problems.

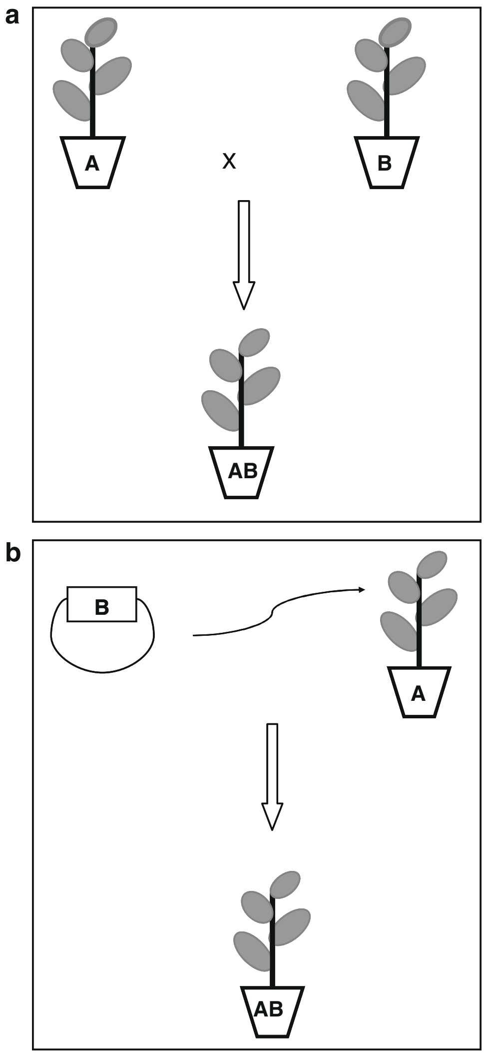


Fig. 26.2 (**a**) Stacking genes A and B by cross-breeding. Plants harbouring transgene A or transgene B are crossed to produce plants with both transgenes. (**b**) Stacking genes A and B by retransformation. A plant harbouring transgene A is retransformed with transgene B

26.2.2 Co-Transformation

Co-transformation is another method used to stack genes in plants, and can be of two categories; single-plasmid co-transformation of linked transgenes (François et al. 2002a; Halpin and Ryan 2004) or multiple-plasmid co-transformation of unlinked transgenes (François et al. 2002a). In the case of single-plasmid co-transformation the

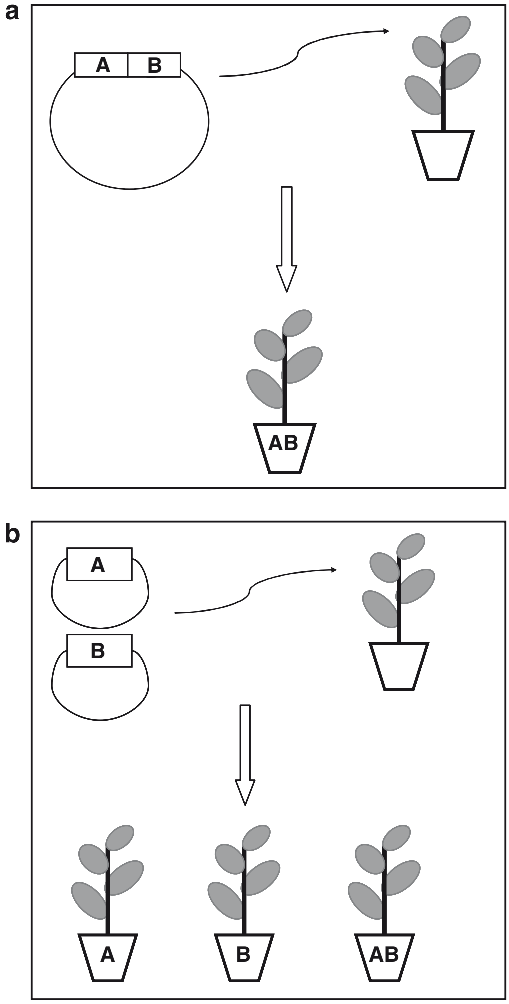


Fig. 26.3 (**a**) Stacking genes A and B by single-plasmid co-transformation. Transgenes A and B are linked on one piece of DNA and transformed together into a plant. (**b**) Stacking genes A and B by multiple-plasmid co-transformation. Transgenes A and B are on different pieces of DNA that are transformed together into a plant

genes to be introduced are linked as a single piece of DNA, with each gene having its own promoter (Fig. 26.3a). Multiple-plasmid co-transformation consists of several plasmids or discrete fragments of DNA (if biolistics are being used), each carrying a different transgene (including a promoter), that are transformed together into a plant (Fig. 26.3b). Single-plasmid co-transformation offers an advantage over multiple-plasmid co-transformation in that integration of both genes together into the same genomic location is ensured as they are linked as a single piece of

DNA (Halpin et al. 2001; François et al. 2002a). However, it is often the case that transgenes co-introduced on multiple plasmids also co-integrate and are therefore inherited together in progeny. As with cross breeding and re-transformation, gene silencing can be a problem (Elmayan and Vaucheret 1996; Voinnet et al. 2003) if the same promoter is used with each transgene to ensure that they are co-ordinately expressed. Use of the same promoter can trigger homology-based silencing (Matzke and Matzke 1995) and therefore it is possible that the introduced gene may not be stably expressed in the long-term (over many plant generations). Nevertheless, multiple-plasmid co-transformation via biolistics has been successful in introducing up to 13 genes simultaneously into plants at one or a few loci (Chen et al. 1998), while six-gene clusters were successfully introduced by single plasmid co-transformation (Goderis et al. 2002; Thomson et al. 2002). Beyond the research level, co-transformation techniques have been successfully used to produce some of the stacked trait GM crops on the market today.

Looking forward to the future, improving methods for single plasmid co-transformation is likely to be an increasingly active research area. The practical problems associated with linking multiple transgenes have already led to the development of a number of new vector systems. Several groups have designed systems where plant transformation vectors harbour an array of rare octa-nucleotide restriction sites and/or homing endonuclease cleavage sites along with unidirectional 'shuttle' vectors with combinations of the same endonuclease sites (Goderis et al. 2002; Thomson et al. 2002). Individual gene cassettes can be constructed in the shuttle vectors then moved over into the plant transformation vector so that artificial gene clusters can be assembled. Other novel systems include Cre/LoxP-based recombination vector systems (Lin et al. 2003), different GATEWAY-based recombination systems (Wakasa et al. 2006; Chen et al. 2006), and several other strategies. Further work is needed to address issues such as improving the delivery of long T-DNAs, overcoming vector size limitations, and determining the stability of long constructs with repeated elements in both bacteria and plant cells (Dafny-Yelin and Tzfira 2007).

26.3 Polycistronic Transgenes

26.3.1 Internal Ribosome Entry Sites (IRES)

A more novel method utilised for expressing multiple transgenes in planta is the use of internal ribosome entry sites (IRES). An IRES is a sequence internal to a mRNA which recruits the ribosome to an initiation codon downstream of the capped 5′-end of the mRNA. IRES sequences, usually of viral origin, can be used in a heterologous context i.e. they can be placed between two transgenes to produce a dicistronic construct (see Fig. 26.4a). Translation of the first open reading frame

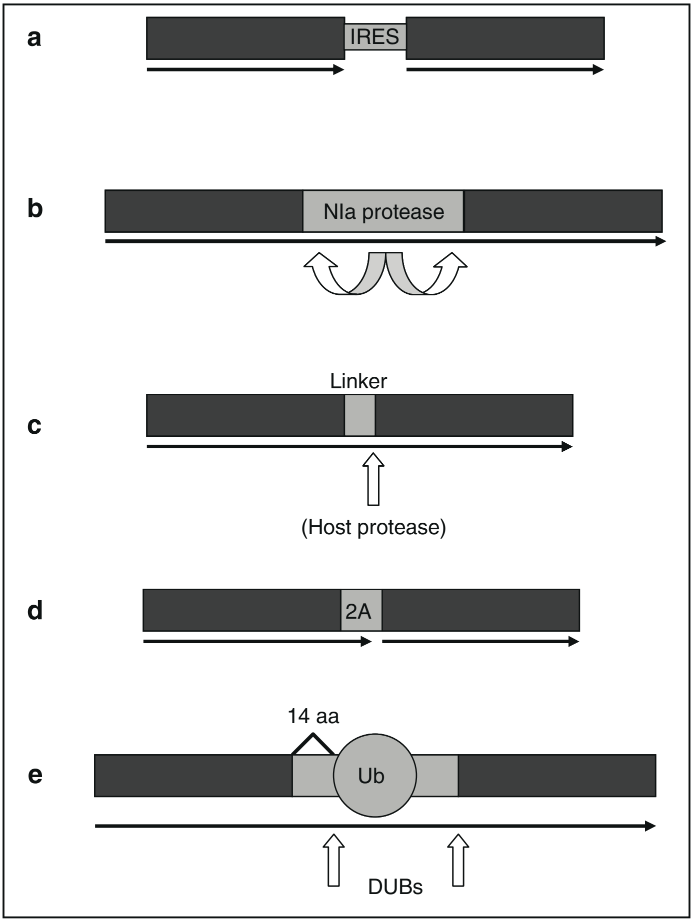


Fig. 26.4 Strategies for gene stacking using single transgene constructs. Black horizontal arrows indicate the primary translation products of the transgene. White arrows indicate sites of post-translational protein processing. (**a**) Polycistronic construct using IRESs. (**b**) Polyprotein construct incorporating NIa protease. (**c**) Polyprotein construct incorporating linker sequence that is cleaved by an endogenous cellular protease. (**d**) Polyprotein construct incorporating FMDV 2A sequence which results in co-translational 'skipping' to yield upstream and downstream protein products. (**d**) Polyprotein construct incorporating ubiquitin that is processed by endogenous deubiquitinating proteases

(ORF) occurs from the first AUG start codon in a cap-dependent manner. Translation of the second open reading frame is initiated from the IRES in a cap-independent manner (François et al. 2002a; Halpin and Ryan 2004). This system is appealing as entire artificial operons can be introduced into plants, and their internal ribosome entry sites recognised (Halpin and Ryan 2004). However, initiation at the IRES can be inefficient, with expression levels of the second open

reading frame being as low as 0.2–1% of the first ORF (François et al. 2002a). No IRES sequences have yet been found in endogenous plant genes (François et al. 2002a), which may explain why re-initiation of translation via an IRES is so poor in planta. However, IRES sequences have been found in plant virus genomes, including Tobamovirus RNA, which suggests that plants may indeed possess a cap-independent translation initiation mechanism (François et al. 2002a). Interestingly, an insect virus (*Rhopalosiphum padi*) IRES was shown to work efficiently in wheat germ lysate (Woolaway et al. 2001). Translation from the construct with the IRES was 17-fold greater than from a construct lacking the IRES, and when an antisense form of the IRES was introduced there was no translation of the second open reading frame. This proved that translation was occurring from the IRES in wheat germ lysate and suggests that the potential does exist for using internal ribosome entry sites in planta.

26.3.2 Chloroplast Transformation

A method of gene stacking that has been used with great success at the research level is stacking via chloroplast transformation ('transplastomic' plants). This method has a major advantage over other methods in that thousands of copies of the chloroplast genome exist within a plant cell, which means that high levels of gene expression can be achieved. The transgenes can be linked together in an operon, high levels of protein expression occur, and epigenetic interference is prevented. As chloroplasts are largely inherited maternally in most crop plants, they are generally not transmitted in pollen, so there is minimal risk of the transgenes being released to the environment. In a study of transplastomic tobacco, only very low levels of paternal plastid transmission could be detected, and the frequency of potentially stable paternal transmission of a transgene was equivalent to 2.86 per million progeny plants (Ruf et al. 2007). It has been proposed that this level of containment is sufficient to enable coexistence of GM and conventional crops but, that other stacking methods or containment methods (e.g. cytoplasmic male sterility) may need to be used in cases where no outcrossing can be tolerated (e.g. for production of specific pharmaceuticals) (Ruf et al. 2007). Plastid transformation is currently only routine in tobacco, although some progress is being made with other crops (see Verma and Daniel 2007 for review). Methods for production of transplastomic plants continue to be improved, for example, several methods for marker-gene excision in transplastomic plants have recently been described (Lutz and Maliga 2007).

Although there are benefits to using chloroplast transformation, there are also some disadvantages. Expression must be tightly regulated, so that the high levels of protein produced are not detrimental to the plants. When three genes for the production of the bioplastic polyhydroxybutyrate (PHB) were linked together to form an operon and transformed into the tobacco chloroplast genome, synthesis of PHB occurred to high levels, but the accumulation of the

plastic prevented the plants from growing properly and caused male sterility (Lössl et al. 2003). These effects were thought to be due to the over-expression of one of the introduced genes. However, this work demonstrates the potential for introducing novel metabolic pathways into plants via chloroplast transformation. Research is being carried out to determine if gene expression could be chemically induced, to mitigate potential detrimental effects when transgenes are expressed throughout development, but this research is still in its early stages (Bock 2007).

26.4 Expression of Multiple Genes from Polyproteins

26.4.1 The NIa Protease

A novel method of expressing multiple transgenes in plants is by expressing transgenes from a self-processing polyprotein. This method is based on the processing of viral genomes as observed in the positive-strand RNA viruses, including the picorna-, poty- and flavi-viruses (Marcos and Beachy 1997). In these, the viral genome is encoded as a single polyprotein which is cleaved in *cis* or in *trans* by proteases within the polyprotein. The Nuclear Inclusion (NIa) protease of the plant potyvirus Tobacco Etch Virus (TEV) is an example of this. Marcos and Beachy (1997) constructed a polypeptide in which the coat protein genes for tobacco mosaic tobamovirus (TMV) and soybean mosaic virus (SMV) were separated by the TEV NIa protease and its cleavage site. When expressed in transgenic tobacco plants, cleavage of the polyprotein to its respective products occurred, however it was found that the proteins were not produced in equimolar quantities, and only accumulated at low levels (Marcos and Beachy 1997). The protein which was N-terminal to the NIa protease was produced in a greater quantity than the protein C-terminal to the NIa protease. It was therefore hypothesised that the nuclear targeting signal within the NIa protease was causing the polyprotein to be directed to the nucleus, resulting in premature termination of translation and unequal quantities of the two protein products (Marcos and Beachy 1997). Further examination of the NIa protease by Ceriani et al. (1998) suggested that disrupting the nuclear targeting of the NIa protease resulted in increased levels of protein expression, however the protein N-terminal to the NIa protease was still present in a greater quantity than the protein C-terminal to the protease. Dasgupta et al. (1998) tried to use transit peptides in polyproteins containing the NIa protease, but found that the protein products did not localise correctly.

Whilst this work shows that the NIa protease can be used to effect cleavage of a polyprotein construct into its separate products (see Fig. 26.4b), the resulting low levels of expression, unequal quantities of proteins and targeting problems prevent it from being a truly useful method of expressing multiple proteins from a single polyprotein construct.

26.4.2 Linker Sequences

It is possible to express multiple genes in plants by separating them with a linker which is then cleaved by a cellular protease. Two examples of linkers processed in this manner are the Kex2 and the AMP (antimicrobial peptide) linkers.

It has been shown that a protease similar to Kex2, a protease required for processing of prohormones and preprotoxins in fungal and animal cells that cleaves at the consensus sequence IGKR↓G (cleavage indicated by ↓), is also present in plant cells in the Golgi apparatus (Kinal et al. 1995; Jiang and Rogers 1999). Kinal et al. 1995 produced transgenic tobacco plants expressing the KP6 preprotoxin from the fungal pathogen Ustilago maydis. Processing of the preprotoxin results in the production and activation of α and β polypeptides. When these two polypeptides were separated by the linker sequence IGKRGKRPR, processing in planta was found to occur at IGKR↓GKRPR and the two active polypeptides were produced. Jiang and Rogers (1999) also demonstrated successful cleavage of a polypeptide where the two coding sequences were separated by a triplicated Kex2 linker sequence (IGKRG IGKRG IGKRG).

A natural polyprotein gene (Ib-AMP) encoding antimicrobial peptides separated by a linker sequence similar to the Kex2 linker, and that is believed to be processed by a protease in the plant endomembrane system, has been identified in Impatiens balsamina (Tailor et al. 1997). The 16 amino acid linker SNAADEVATPEDVEPG has doublets of negatively charged amino acids (DE, ED) which are hypothesised to be the sites of cleavage (Tailor et al. 1997; François et al. 2002b). François et al 2002b constructed a heterologous polypeptide by placing the 16 amino acid linker sequence between two different antimicrobial peptides (DmAMP1 from Dahlia merckii and RsAFP2 from Raphanus sativus) in a single open reading frame. Cleavage of the polyprotein occurred within the linker sequence, although the exact position of the cleavage was not determined. The cleavage resulted in the production of DmAMP1 with an additional serine residue attached to its carboxy-terminus and RsAFP2 with five additional amino acids present at its N-terminus (DVEPG). When the order of the genes in the polypeptide was reversed, similar cleavage results were observed. However, François et al 2002b discovered that cleavage occurred both directly after the RsAFP2 domain and after the adjacent serine residue to yield the products RsAFP2 and RsAFP2 + Ser. This data suggested that cleavage of the linker sequence occurs independently of the genes N- and C-terminal to it.

The protease responsible for cleaving the Ib-AMP linker sequence remains unknown. It is hypothesized that an endoprotease may be responsible, perhaps in combination with another endo- or exoprotease. It was proposed by François et al 2002b that cleavage at the linker sequence first occurs by a specific endoprotease, the linker is then subsequently "trimmed" by an exoprotease.

Both the Kex2 and Ib-AMP linker sequences have been used successfully to demonstrate expression of multiple genes from a single polyprotein construct in planta (see Fig. 26.4c). However, there are limitations to using these methods. These linker sequences would only be useful if both protein products were to be targeted to the endomembrane system where the proteases are thought to reside.

Also, since the proteases responsible for cleavage remain to be conclusively identified it is possible that they may not be present in all plant tissues or species, so cleavage could be tissue- or species-dependent.

26.4.3 FMDV 2A

Dissociation of a polyprotein resulting in the expression of multiple proteins in plants has been demonstrated using the 2A peptide from Foot and Mouth Disease Virus (FMDV) (Halpin et al. 1999). In this method the 20 amino acid 2A peptide is placed between two genes, forming a single open reading frame. Upon translation dissociation occurs at the C-terminus of the 2A peptide, yielding the respective protein products (Ryan and Drew 1994; Halpin et al. 1999). The mechanism by which cleavage occurs is still debated, though it is proposed that during translation of the polyprotein, the conformation of the 2A peptide prevents a peptide bond forming between a glycine and proline residue at the conserved 2A cleavage site. This ribosomal 'skip' results in the first protein being released from the ribosome, while translation of the second protein continues (Donnelly et al. 2001a). The system ensures co-ordinate production of both proteins in broadly similar amounts (Halpin et al. 2001; François et al. 2002b). Genes containing transit peptides or signal sequences have been used in conjunction with 2A, and the proteins have been found to target correctly to their subcellular locations: chloroplast and ER respectively (Halpin et al. 2001; El Amrani et al. 2004).

There are many advantages of using the 2A peptide to introduce multiple genes into plants. The sequence is short and multiple genes can be introduced in a single open reading frame (see Fig. 26.4d). There is no need to rely on the activity of cellular proteases to effect cleavage, as the 'cleavage activity' is inherent to the peptide. Problems of homology-based silencing are reduced as only a single promoter is used (Halpin et al. 1999; François et al. 2002a) and the method can be used in all plant tissues, regardless of the developmental stage (Halpin et al. 2001). Many other '2A-like' sequences have been identified, mainly from other viruses (Donnelly et al. 2001b, Osborn et al. 2005) which may also cleave in planta.

However there are also potential limitations to using the 2A peptide. Upon dissociation of the polypeptide, 19 amino acids from the 2A sequence remain attached to the C-terminus of the first protein. It is possible that this might interfere with protein folding and function or cause protein mis-targeting if the protein has a C-terminal targeting sequence (Halpin et al. 1999). To date however, these problems have not arisen when such 2A-containing polyproteins have been used in plant or animal systems (Halpin et al. 1999, 2001; Szymczak et al. 2004). The single proline residue from the 2A peptide that remains attached to the N-terminus of the second protein does not appear to cause problems either and, according to the N-end rule, this residue should serve to promote protein stability (Halpin et al. 1999; Halpin and Ryan 2004).

Ma and Mitra (2002) successfully demonstrated the cleavage of a polyprotein containing three genes separated by the 2A peptide. The polypeptide was successfully processed at both the first and second 2As, and the two 2A sequences acted independently of each other. El Amrani et al. (2004) expressed genes containing chloroplast transit peptides using the 2A system. They found that after dissociation at 2A, the proteins were post-translationally targeted to the chloroplasts, and the transit peptide was removed. This therefore demonstrated that following rapid co-translational dissociation, proteins could be targeted to their appropriate sub-cellular location, and the presence of the excess residues from the 2A sequence did not interfere with this process.

26.4.4 Ubiquitin Vectors

A more recent approach to gene stacking in planta involves the use of ubiquitin (Ub)-based vectors. Similarly to FMDV 2A, these vectors allow co-ordinate expression of multiple proteins from a single chimeric transgene. In vivo, ubiquitin is synthesized either attached to the N-terminus of proteins or as polyubiquitin, and deubiquitinating enzymes (DUBS) process the ubiquitin fusions into their respective, functional components. The deubiquitinating enzymes are specific for ubiquitin and can remove most proteins joined to the ubiquitin C-terminus. Thus, in yeast, fusion constructs where two proteins are connected in-frame with a single Ub sequence have been shown to be cleaved after the last residue of Ub to yield equimolar amounts of the two proteins, with the anterior one carrying the C-terminal Ub unit (Levy et al. 1996; Suzuki and Varshavsky 1999).

This strategy has recently been slightly refined and used in plants. Processing of a chimeric polyprotein expressed from a ubiquitin-based vector was demonstrated by Walker and Vierstra (2007), when they used the method to express luciferase (LUC) and β-glucoronidase (GUS) in stably transformed tobacco plants. Their construct consisted of luciferase followed by 14 C-terminal amino acids from ubiquitin (Ub linker), joined to a full-length ubiquitin and then to GUS (LUC-14aa-Ub-GUS). Processing of the polyprotein was expected to result in the production of luciferase (with the 14 amino acid Ub linker attached to it) and GUS, as this had previously been observed in studies with yeast. Western blotting of protein extract derived from the plants showed that, as expected, GUS was produced, but the resulting LUC was closer in size to native LUC than to LUC with fused Ub linker. The authors suggest that this indicates cleavage by a tobacco DUB at the junction between LUC and the start of the Ub linker sequence. Subsequent over-exposure of the blot failed to show the presence of any unprocessed polyprotein, indicating that the efficiency of polyprotein processing by DUBs was very high. They further tried to improve this technique by reducing the 14 amino acid Ub linker to shorter lengths. They still achieved complete processing with a 6 amino acid linker but further reduction to 4 amino acids reduced the efficiency of processing (Walker and Vierstra 2007).

Whilst this system is not without fault there are several advantages to using this method for gene stacking in plants. The deubiquitinating enzymes are endogenous to the plant meaning that foreign proteases do not have to be introduced to ensure polyprotein processing occurs. Upon processing only14 amino acids remain attached to the protein anterior to the ubiquitin (see Fig. 26.4e). This can be reduced to 6 amino acids without affecting cleavage, or less if only partial processing is deemed acceptable. The production of protein products with no amino acid attachments would be particularly useful to ensure correct functioning and targeting of the products, and for gaining regulatory approval. Another benefit of the Ub-based vector system is that there appears to be little or no unprocessed polyprotein produced. This is possibly due to plants having a greater concentration or more active DUBs (Walker and Vierstra 2007). It is thought that the presence of the ubiquitin moiety should not interfere with protein targeting; therefore it should be possible to target proteins to different subcellular locations. Walker and Vierstra (2007) suggest that more than three proteins could be expressed using this system and that the stoichiometry of the proteins could be controlled, for example repeating one of the coding sequences to create a 2:1 ratio of different proteins. Theoretically any number of proteins could be expressed using this vector system however as with several other systems, the length of the transcribed mRNA may become a limiting factor. If it is too long it may be unstable and prevent transcription of the full polyprotein. It was also noted by Walker and Vierstra (2007) that the quantities of protein produced in planta using the polyprotein cassettes was less than that produced by non-fused LUC and GUS transgenes using the same promoter.

26.4.5 Suppression of Multiple Genes via Chimeric Transgenes

Whilst transgenes can be introduced into plants to confer novel traits or increase expression of existing genes, they can also be used to co-ordinately suppress expression of existing genes within a plant. Abbott et al. (2002) demonstrated that it was possible to suppress the expression of three lignin genes by creating a construct where partial sense sequences for the three genes were fused together and expressed from the 35S CaMV promoter. When the transgenic plants were examined it was found that levels of all three genes were significantly below that of the wild-type plants and were similarly suppressed, indicating that co-ordinate down-regulation had occurred. When Abbott et al. (2002) tried to obtain double antisense plants by crossing they found that, while some of the progeny had reduced levels of gene expression, many had levels which were considerably higher than expected, and the levels varied significantly between plants. It was thought that silencing was triggered by crossing two plants containing genes with the same promoter. Therefore, it would appear that for co-ordinate suppression of genes it is best to construct a chimeric transgene. This method also has several other advantages over crossing plants to down-regulate gene expression. It is very quick, taking only one generation to produce plants with multiple genes suppressed. Further breeding to produce homozygous transgenic plants is

easier than it is using other methods where the combined transgenes may resegregate. Gene suppression is coordinate due to the partial sense sequences being linked together as a single transcript and this also means that the promoter is not duplicated, so homology based silencing should be minimal (Abbott et al. 2002).

26.5 Gene Stacking Using Mini-Chromosomes

The production of minichromosomes is a very recently developed method, still being actively refined, which could be extremely useful for gene stacking in plants. Autonomous, independently-segregating chromosome-based vectors could offer a stable platform for introducing and maintaining large numbers of novel genes, even entire new metabolic pathways, into plants. Although already successfully used in mammals and yeast, artificial chromosome engineering in plants has only recently made significant progress with several reports of the production of transgene-expressing minichromosomes in maize (Carlson et al. 2007; Yu et al. 2006, 2007).

Yu et al. (2007) succeeded in creating minichromosomes from endogenous chromosomes in maize by introducing telomere sequences that cause chromosomal truncation at the insertion site. This chromosome engineering was achieved by Agrobacterium-mediated transformation of maize embryos with constructs containing a selectable marker, a site-specific recombination cassette and telomere repeats. Upon integration into chromosomes, the telomere-containing transgene capped the broken chromosome and caused formation of a new telomere, hence a truncated ‘minichromosome’ was formed. By including site specific recombination sites within the transforming DNA, other transgenes could be introduced into the minichromosome at a later date. Expression of a minichromosome-harboured GUS transgene was detected in all of the tissues examined, including the roots, shoots, leaves and endosperm. An advantage of this minichromosome system is that, by targeting the supernumerary B chromosome, dosage manipulations can be carried out due to the nondisjunction properties of B minichromosomes that occur during mitosis. This means that in the presence of normal B chromosomes, male parents can transmit two copies of the truncated chromosome. Therefore by selective crossing, multiple minichromosomes could potentially be accumulated in one plant. The minichromosomes produced by Yu et al. could be replicated through meiosis and transmitted, with varying efficiencies, to the next generation. However, further work is needed to fully understand the behaviour of these minichromosomes through meiosis and to improve the efficiency and stability of transmission.

An alternative method for chromosome engineering involves de novo assembly of circular minichromosomes by transforming maize embryogenic tissue with autonomous vectors containing up to 190kb of maize centromeric repeat sequences (Carlson et al. 2007). After selection for the marker gene present within the vector, fluorescence in situ hybridization (FISH) revealed some events where the vector had integrated into chromosomes but other events where the FISH signal was clearly distinct from the DAPI-stained host chromosomes, suggesting autonomous

maintenance. One of these autonomous minichromosomes was further characterized in detail and found to be efficiently transmitted though both mitotic and meiotic cell divisions through at least four generations (Carlson et al. 2007).

The advent of minichromosome technology for plants, while still nascent and deserving of much further research, offers the potential for greatly simplifying the production of stacked trait transgenics while also increasing their speed-to-market and has already attracted significant interest from industry.

26.6 Conclusions

An increasing 'toolkit' of technologies is developing for stacking useful genes and traits in transgenic plants. Some of these techniques have already been used in the commercial production of stacked trait GM crops, and an increasingly rapid pace of advance in the introduction of such stacked gene products is expected. Further refining and widening of the options available for efficient multiple gene manipulation in plants is a major challenge for the future if we are to develop more sophisticated and cleaner transgenic technologies which will allow the full opportunities for GM crops to be realized.

References

Abbott JC, Barakate A et al. (2002) Simultaneous suppression of multiple genes by single transgenes. Down-regulation of three unrelated lignin biosynthetic genes in tobacco. Plant Physiol 128:844–853

Bizily SP, Rugh CL et al. (2000) Phytodetoxification of hazardous organomercurials by genetically engineered plants. Nat Biotechnol 18:213–217

Bock R (2007) Plastid biotechnology: prospects for herbicide and insect resistance, metabolic engineering and molecular farming. Curr Opin Biotechnol 18:100–106

Ceriani MF, Marcos JF et al. (1998) Simultaneous accumulation of multiple viral coat proteins from a TEV-NIa based expression vector. Plant Mol Biol 36: 239–248

Carlson SR, Rudgers GW et al. (2007) Meiotic transmission of an in vitro–assembled autonomous maize minichromosome. PLoS Genet 3(10): e179. doi:10.1371/journal.pgen.0030179

Chen LL, Marmey P et al. (1998) Expression and inheritance of multiple transgenes in rice plants. Nat Biotechnol 16:1060–1064

Chen QJ, Zhou HM et al. (2006) A gateway-based platform for multigene plant transformation. Plant Mol Biol 62:927–936

Dafny-Yelin M and Tzfira T (2007) delivery of multiple transgenes to plant cells. Plant Physiol 145:1118–1128

Dasgupta S, Collins GB et al. (1998) Co-ordinated expression of multiple enzymes in different subcellular compartments of plants. Plant J 16:107–116

Donnelly MLL, Luke G et al. (2001a) Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. J Gen Virol 82:1013–1025

Donnelly MLL, Hughes LE et al. (2001b) The 'cleavage' activities of Foot-and-MouthDisease Virus 2A site-directed mutants and naturally occurring '2A-like' sequences. J Gen Virol 78:13–21

El Amrani A, Barakate A et al. (2004) Coordinate expression and independent subcellular targeting of multiple proteins from a single transgene. Plant Physiol 135:16–24
Elmayan T and Vaucheret H (1996) Expression of single copies of a strongly expressed 35S transgene can be silenced post-transcriptionally. Plant J 9:787–797
François IEJA, Broekaert WF et al. (2002a) Different approaches for multi-transgene-stacking in plants. Plant Sci 163:281–295
François IEJA, De Bolle MFC et al. (2002b) Processing in transgenic *Arabidopsis thaliana* plants of polyproteins with linker peptide variants derived from the *Impatiens balsamina* antimicrobial polyprotein precursor. Plant Physiol Biochem 40:871–879
Goderis IJWM, De Bolle MFC et al. (2002) A set of modular plant transformation vectors allowing flexible insertion of up to six expression units. Plant Mol Biol 50:17–27
Halpin C (2005) Gene stacking in transgenic plants Ð the challenge for 21st century plant biotechnology. Plant Biotechnol J 3:141–155
Halpin C and Ryan M (2004) Redirecting metabolism by co-ordinate manipulation of multiple genes. Metabolic engineering in the post genomic era. WHV Kholodenko, Horizon Bioscience, pp 377–408
Halpin C, Cooke SE et al. (1999) Self-processing 2A-polyproteins – a system for co-ordinate expression of multiple proteins in transgenic plants. Plant J 17:453–459
Halpin C, Barakate A et al. (2001) Enabling technologies for manipulating multiple genes on complex pathways. Plant Mol Biol 47:295–310
James, C (1997) Global status of transgenic crops: 1997. ISAAA Briefs No. 5. ISAAA: Ithaca, NY
James C (1998) Global review of commercialized transgenic crops: 1998. ISAAA Briefs No.8. ISAAA: Ithaca, NY
James, C (1999) Global review of commercialized transgenic crops: 1999. ISAAA Briefs No.12: Preview. ISAAA: Ithaca, NY
James, C (2000) Global status of commercialized transgenic crops: 1999. ISAAA Briefs No.17. ISAAA: Ithaca, NY
James, C (2001) Global status of commercialized transgenic crops: 2000. ISAAA Briefs No. 23. ISAAA: Ithaca, NY
James, C (2002) Global status of commercialized transgenic crops: 2002. ISAAA Briefs No. 26. ISAAA: Ithaca, NY
James, C (2003) Global status of commercialized biotech/GM crops: 2007. ISAAA Brief No. 30. ISAAA: Ithaca, NY
James, C (2004) Preview: global status of commercialized transgenic crops: 2004. ISAAA Briefs No. 32. ISAAA: Ithaca, NY
James, C (2005) Global status of commercialized biotech/GM crops: 2007. ISAAA Brief No. 34. ISAAA: Ithaca, NY
James, C (2006) Global status of commercialized biotech/GM crops: 2007. ISAAA Brief No. 35. ISAAA: Ithaca, NY
James, C (2007) Preview: global status of commercialized transgenic crops: 2007. ISAAA Brief No. 37. ISAAA: Ithaca, NY
Jiang L and Rogers JC (1999) Functional analysis of a Golgi-localized Kex2p-like protease in tobacco suspension culture cells. Plant J 18:23–32
Kinal H, Park C et al. (1995) Processing and secretion of a virally encoded antifungal toxin in transgenic tobacco plants: evidence for a Kex2p pathway in plants. Plant Cell 7:677–688
Lapierre C, Pollet B et al. (1999) Structural alterations of lignins in transgenic poplars wit depressed cinnamyl alcohol dehydrogenase or caffeic acid O-methyltransferase activity have an opposite impact on the efficiency of industrial kraft pulping. Plant Physiol 119:153–163
Levy F, Johnsson N.et al. (1996) Using ubiquitin to follow the metabolic fate of a protein. Proc Natl Acad Sci U S A 93:4907–4912
Lin L, Liu YG et al. (2003) Efficient linking and transfer of multiple genes by a multigene assembly and transformation vector system. Proc Natl Acad Sci U S A 100:5962–5967
Lössl A, Eibl C et al. (2003) Polyester synthesis in transplastomic tobacco (*Nicotiana tabacum* L.): significant contents of polyhydroxybutyrate are associated with growth reduction. Plant Cell Rep 21:891–899

Lutz A and Maliga P (2007) Construction of marker-free transplastomic plants. Curr Opin Biotechnol 18:107–114

Ma C and Mitra A (2002) Expressing multiple genes in a single open reading frame with the 2A region of Foot-and-Mouth Disease Virus as a linker. Mol Breed 9:191–199

Ma JKC, Hiatt A et al. (1995) Generation and assembly of secretory antibodies in plants. Science 268:716–719

Marcos JF and Beachy RN (1997) Transgenic accumulation of two plant virus coat proteins on a single self-processing polypeptide. J Gen Virol 78:1771–1778

Matzke MA and Matzke AJM (1995) How and why do plants inactivate homologous (*trans*) genes? Plant Physiol 107:679–685

Osborn MJ, Panoskaltis-Mortari A et al. (2005) A Picornaviral 2A-like sequence-based tricistronic vector allowing for high-level therapeutic gene expression coupled to a dual-reporter system. Mol Ther 12:569–574

Rosati C, Simoneau P et al. (2003) Engineering of flower color in forsythia by expression of two independently-transformed dihydroflavonol 4-reductase and anthocyanidin synthase genes of the flavonoid pathway. Mol Breed 12:197–208

Ruf S, Karcher D et al. (2007) Determining the transgene containment level provided by chloroplast transformation. Proc Natl Acad Sci U S A 104: 6998–7002

Ryan MD and Drew J (1994) Foot-and-Mouth Disease Virus 2A oligopeptide mediated cleavage of an artificial polyprotein. EMBO J 13:928–933

Suzuki T and Varshavsky A (1999) Degradation signals in the lysine–asparagine sequence space. EMBO J 18:6017–6026

Szymczak AL, Workman CJ et al. (2004) Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. Nat Biotechnol 22:589–594

Tailor RH, Acland DP et al. (1997) A novel family of small cysteine-rich antimicrobial peptides from seed of Impatiens balsamina is derived from a single precursor protein. J Biol Chem 272:24480–24487

Thomson JM, Lafayette PR et al. (2002) Artificial gene-clusters engineered into plants using a vector system based on intron- and intein-encoded endonucleases. In Vitro Cell Dev Biol - Plant 38:537–542

Verma D and Daniell H (2007) Chloroplast vector systems for biotechnology applications. Plant Physiol 145:1129–1143

Voinnet O, Rivas S et al. (2003) An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. Plant J 33:949–56

Wakasa Y, Yasuda H et al. (2006) High accumulation of bioactive peptide in transgenic rice seeds by expression of introduced multiple genes. Plant Biotechnol J 4:499–510

Walker JM, Vierstra RD (2007) A ubiquitin-based vector for the co-ordinated synthesis of multiple proteins in plants. Plant Biotechnol J 5:413–421

Woolaway KE, Konstantinos L et al. (2001) The 5' untranslated region of *Rhopalosiphum padi* Virus contains an internal ribosome entry site which functions effectively in mammalian, plant and insect translation systems. J Virol 75:10244–10249

Yu W, Lamb JC et al. (2006) Telomere-mediated chromosomal truncation in maize. Proc Natl Acad Sci U S A 103:173331–173336

Yu W, Han F et al. (2007) Construction and behaviour of engineered minichromosomes in maize. Proc Natl Acad Sci U S A 104:8924–8929

Chapter 27
Gene Silencing

Sunee Kertbundit, Miloslav Juříček, and Timothy C. Hall

Abstract Early experiments in which a gene was transferred from one species to another were very exciting but were often disappointing in that the gene expression seen in the first generation was not carried through to subsequent generations. It also became clear that this silencing was not a result of incomplete or erroneous translation. A major breakthrough was the work of Dougherty and colleagues who showed that aberrant RNA led to lengthy but often incomplete or rearranged RNA species. By 1998, it was recognized both in animal and plant systems that rearrangements of the transgene that allowed anti-parallel RNA strands to be generated stimulated a templated degradation system (RISC complex) that includes specific small RNAs together with Argonaute proteins and Dicer-like proteins that cleave the targeted RNA into characteristic 21–25 nt degradation products. In addition to induction by anti-parallel RNAs, RNAi degradation systems can be induced by virus replication intermediates as in natural host defense mechanisms. By 2001, another kind of small RNA involved in regulation of gene expression was discovered and termed microRNA. These endogenous ~21 nt small RNAs play several roles in gene regulatory networks and often target other regulatory genes. Artificial small RNAs can also be designed to inactivate target genes to generate the mutant phenotypes which are useful for crop improvement. Thus, whereas the instability of transgene expression is seen as an undesirable feature, correctly used, similar strategies are proving powerful in debilitating many pathogens for which no protection has been available from classical sources.

S. Kertbundit and M. Juříček
Institute of Experimental Botany, Academy of Sciences of the Czech Republic, Na Pernikářce 15, CZ – 160 00, Prague 6, Czech Republic

S. Kertbundit
Department of Plant Science, Faculty of Science, Mahidol University, Rama VI Road, Bangkok, 10400, Thailand

T.C. Hall (✉)
Institute of Developmental and Molecular Biology, Department of Biology, Texas A&M University, College Station, TX 77843-3155, USA
e-mail: tim@idmb.tamu.edu

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_27,

27.1 Introduction

Transgenic plants were first developed in 1983 by four independent research groups (Bevan et al. 1983; Fraley et al. 1983; Herrera-Estrella et al. 1983; Murai et al. 1983). These achievements opened new horizons for plant research and crop improvement. Numerous genes have been introduced into crop plants to develop new varieties that have desired traits such as transgenic corn and cotton expressing the *Bacillus thuringiensis* crystal toxin that can kill European corn borer, the bollworm and other lepidopteran insects. However a major problem soon became apparent: many of the transgenes were silenced, i.e. they did not express the encoded gene product as expected.

At first, transgene silencing was thought to be similar to the situation for Drosophila *Polycomb* genes that become heterochromatinized and their DNA methylated. However, as experimental data accumulated, it became apparent that several events are involved in gene silencing including RNA silencing, epigenetic events, remodeling of chromatin structure and others. Intriguingly, gene silencing is now considered to be an important mechanism by crop breeders, as it allows the inactivation of certain genes without the need for lengthy breeding. In some cases, the inactivation of a particular gene could result in an improved quality crop. In this chapter, we review current concepts concerning mechanisms of gene silencing and their use for crop improvement, with a special focus on rice.

27.2 Co-suppression and Gene Silencing

Gene silencing of transgenic DNA in plants was initially observed in petunia (*Petunia hybrida*) transformed with a chalcone synthase A (*chsA*) gene and termed homology-dependent gene silencing (HDGS) (van der Krol et al. 1990) or co-suppression (Napoli et al. 1990). Chalcone synthase (CHS) is a key enzyme in the anthocyanin biosynthesis pathway and the introduction of an extra copy of *chsA* into a purple-flowered petunia line was expected to yield plants bearing flowers of more intense color. In practice, white-flowered progeny were obtained and it was subsequently discovered that the mRNAs from both the host *chsA* gene and homologous transgenes were degraded ("co-suppressed") following transcription in the nucleus, accounting for the lack of CHS expression. This co-suppression phenomenon has now been observed in many other organisms, notably in *Neurospora crassa* (Cogoni and Macino 1999), *Drosophila melanogaster* (Pal-Bhadra et al. 1997, 1999) and *Caenorhabditis elegans* (Gaudet et al. 1996; Jones and Schedl 1995).

Co-suppression, along with various forms of viral gene silencing (Baulcombe 1999; English et al. 1996; Lindbo et al. 1993), is now recognized as a post-transcriptional gene silencing (PTGS) event (van Blokland et al. 1994). The term "co-suppression" is still used sometimes to describe a silencing strategy that is based on integration of additional copies of a gene.

27.3 Transgene Silencing

Several studies have documented a higher incidence of insertion of multiple copies of complex, rearranged transgenes from direct DNA transfer methods than from *Agrobacterium*-mediated transformation (Cheng et al. 1998; Hiei et al. 1994). Further, a high incidence of silencing has been observed when the transgene is rearranged or when multiple copies are present as tandem or interspersed direct or inverted repeats (Assaad et al. 1993; Dorer and Henikoff 1997; Garrick et al. 1998; Rossignol and Faugeron 1994). The presence of multiple copies provides additional opportunities for ectopic interactions between transgene sequences and has been proposed to cause activation of silencing (Kumpatla and Hall 1998b).

Gene silencing can occur at the transcriptional (TGS) or at the post-transcriptional (PTGS) level. Useful generalizations, to which there are many exceptions, are that: whereas in TGS the promoter is methylated and no mRNA is transcribed from the transgene, in PTGS, although the coding sequences are methylated, the mRNA is transcribed but is rapidly degraded (Kooter et al. 1999; Matzke and Matzke 1995; Vaucheret et al. 1998). Superficially, an inability to detect the target gene mRNA (either by Northern blot, nuclear run-on or RNase protection assays) appears to provide *de facto* evidence for TGS. However, it is now recognized that the initiating step could be a short-lived PTGS event that sets up a more long-lived TGS system. TGS is both mitotically and meiotically heritable, whereas PTGS is reversed following meiosis (Depicker and Montagu 1997; Stam et al. 1998). Other early conjectures were that TGS and PTGS are involved both in normal gene regulation processes and in host defense responses to extragenomic and intragenomic parasites (Bestor 1998; Kumpatla et al. 1998; Matzke and Matzke 1998). Nevertheless, it was recognized that silencing of a specific transgene could result from a combination of both TGS and PTGS effects (Vaucheret and Fagard 2001).

27.4 Gene Silencing in Rice

Rice was the first monocot crop plant for which routine molecular transformation was established. Of the several different approaches that have been used, early successful transformations were by electroporation (Zhang et al. 1988) or polyethylene glycol-mediated transformation of protoplasts (Li et al. 1990). Protoplasts are very frail and regeneration requires extensive tissue culture, the outcome being a high frequency of somaclonal variant plants that are often sterile. The biolistic or particle gun bombardment approach (Cao et al. 1992; Klein et al. 1987) dramatically reduced these events and *Agrobacterium*-mediated transformation of embryogenic rice calli (Hiei et al. 1994) opened the door for the biotechnological development of this vital crop.

A comprehensive study of transgene expression in rice was undertaken by the Hall laboratory using biolistic transformation with *Btt cryIIIA*, a synthetic gene that codes for a putatively insecticidal protein (CRYIIIA), under the control of the CaMV *35S* promoter; and the *bar* gene that confers resistance to the herbicide

bialaphos (BAR), driven by the maize ubiquitin1 promoter (*mUbiI*). Aberrant segregation patterns for both CRYIIIA expression (Kumpatla and Hall 1999) and BAR resistance (Kumpatla and Hall 1998a) were observed in progeny as a result of silencing. Analysis of coding region methylation indicated that silencing of the *bar* gene was methylation-based. Nuclear run-on assays confirmed that *bar* gene silencing was at the transcriptional level, although it now seems likely that initial steps in silencing may have resulted from PTGS events triggered by the rearrangements present in the integrated construct. Interestingly, R2 and R3 progeny of plants that were not silenced in the R1 or R2 generations were frequently silenced (Kumpatla and Hall 1998b). With one possible exception, no spontaneous re-activation of expression was observed, but germination of seeds in solution containing 5-azacytidine or Trichostatin A typically restored expression for up to one month (Kumpatla and Hall 1998a). The apparently random silencing of isogenic seed lines and the universal re-silencing of re-activated plants are highly undesirable events for crop plants, making an understanding of the causes of silencing an important applied goal.

In order to investigate the relationship between copy number and silencing (HDGS) we (Yang et al. 2005) made further use of the rice line containing the silenced *mUbi1* promoter. The line JKA 52 was established as being homozygous at the molecular level by genomic DNA blots and genetically in that all offspring were silenced. The rationale for the experiment was that, if a gene bearing an identical promoter were introduced into a genomic background in which that *Agrobacterium*-mediated promoter is silenced, on the basis of HDGS, little to no resistance was expected. A *mUbi1-GUS* construct was used to supertransform into the *bar*-silenced line JKA 52. All twelve independent supertransformants expressed GUS but none were herbicide resistant, i.e. the resident *mUbi1* promoter was not reactivated and the incoming *mUbi1* promoter was not silenced.

Similarly, the inheritance and expression of a transgene locus consisting of multiple copies of a rice chitinase gene under the control of the *CaMV 35S* promoter was studied in the T3 and T4 generations of plants derived from PEG-mediated transformation of rice protoplasts with chitinase and hygromycin phosphotransferase (*hpt*) genes (Chareonpornwattana et al. 1999). Here again, it was concluded that the critical event for the induction of silencing was not copy number but expression level. Once the silent phenotype was established, it was stable and meiotically heritable. The silencing of chitinase was shown by nuclear run-on experiments to be at the transcriptional level in the lines that were studied, but the endogenous chitinase gene was not silenced.

The findings of (Yang et al. 2005) and (Chareonpornwattana et al. 1999) are in accordance with current understanding that aberrant transcripts, especially those leading to the production of double-stranded RNA are potent silencing signals. Indeed, the JKA lines used in the studies of Yang et al. 2005 contained a *35S-hyg* gene in addition to the *mUbi1-bar* gene. Earlier work (Kumpatla and Hall 1999) had shown the presence of multiple rearrangements of the *35S* promoter and these may have resulted in RNAi-induced silencing of *hpt* whereas no rearranged transcripts or silencing of the incoming *mubi1-bar* gene were detected.

Molecular characterization of the rice obtained from biolistic transformation with constructs containing *gus*, *hpt* and *bar* coding regions, each driven by a 35S promoter showed that the transgene was typically found to be inserted at one locus and was present in 1–9 copies of rearranged or truncated sequence (Kohli et al. 1999). The increase in transgene copy number did not always lead to a concomitant decrease in expression levels or silencing but the integrity of integrated transgenes was a major factor in the onset of silencing. They observed that the presence of truncated sequences of transgenes capable of generating incomplete transcripts, resulting in aberrant RNA species, may be responsible for silencing.

Silencing of GUS driven by a barley aleurone-specific lipid transfer protein (*ltp*) promoter in transgenic rice plants obtained from protoplast electroporation (Morino et al. 1999) caused spotted GUS patterns on the grain surface. Various patterns of silencing were observed in the R2 generation and were inherited for five generations. Differences in the frequencies of silencing were attributed to developmental and environmental growth conditions. Two transgene loci were found, one of which carried a rearranged copy containing part of the *gus* gene in the antisense orientation. The RNA analyses showed that this rearranged transgene generated polyadenylated *gus* transcripts. The authors suggested that partial antisense RNAs generated from the rearranged transgene locus or truncated RNAs derived from them interact with normal RNA derived from the intact transgene and that this interaction leads to silencing. A similar example of grain patterning was observed for GUS expression driven by the bean *phas* promoter in transgenic *Medicago truncatula* seeds (Zhou et al. 2005). These examples of RNAi-derived silencing show that insight to agronomic characteristics such as grain patterning can be gained from basic studies.

27.5 Post-transcriptional Gene Silencing and Virus Resistance

PTGS in plants is analogous to co-suppression, RNA interference (RNAi) in *C. elegans* (Fire et al. 1998) and quelling in *Neurospora* (Cogoni and Macino 1999). It is implicated as a natural defense mechanism to counter viruses and transposable elements (Baulcombe 2004, 1996; Brodersen and Voinnet 2006; Hamilton and Baulcombe 1999; Montgomery and Fire 1998; Vaucheret 2006). PTGS against virus infection can also be triggered in transgenic plants containing virus sequences. In the beginning, it was understood that the pathogen-derived resistance (PDR), an alternative approach for crop protection by introducing a genome segment of the pathogen into the plant genome (Sanford and Johnston 1985), was mediated by the protein produced from the viral transgene (Brederode et al. 1995; Lapidot et al. 1993; Powell-Abel et al. 1986). However, it was later shown that transformation of both a full-length form of the tobacco etch virus (TEV) coat protein or a truncated form at the N terminus of the TEV coat protein into tobacco plants showed high virus resistance (Lindbo and Dougherty 1992). The nuclear run-off assays showed that introduced TEV sequences were actively transcribed, but the transgene transcript failed to accumulate. From these results, they proposed that the TEV resistant

was mediated by RNA degradation in the cytoplasm at the post transcriptional level (Lindbo et al. 1993). Subsequently studies have linked PTGS and virus resistance in transgenic plants containing virus-derived transgenes (English et al. 1996; Goodwin et al. 1996; Jacobs et al. 1997; Tanzer et al. 1997). This type of virus resistance became known as RNA-mediated virus resistance (RMVR). It can confer high level of resistance that was not easily overcome by a high inoculum dose (Lomonossoff 1995) and specific to closely related viruses (Baulcombe 1996; Lomonossoff 1995; Prins and Goldbach 1996).

In rice, transformation of a chimeric 35S CaMV-viral RNA-dependent RNA polymerase (RdRP) of rice yellow mottle virus (RYMV) into immature rice embryos of African lowland indica varieties by particle gun bombardment generated transgenic lines resistant to RYMV (Pinto et al. 1999). In the most extreme examples there was complete suppression of virus multiplication. Resistance was stable over at least three generations. The nuclear run off analysis indicated that the resistance derives from an RNA-based mechanism associated with PTGS.

27.6 RNA Mediated Gene Silencing

Several studies on PTGS revealed the presence of small RNAs that are either truncated RNAs generated by premature termination of the transgene transcript or degradation of full-size RNA in silenced plants (English et al. 1996; Goodwin et al. 1996; Jacobs et al. 1997; Metzlaff et al. 1997; Tanzer et al. 1997). Potentially, these small RNAs can trigger degradation of all transcripts corresponding to the targeted gene (Baulcombe and English 1996; Depicker and Montagu 1997). The basic pathway of RNA-mediated silencing (Fig. 27.1) is similar from higher organisms such as animals and plants to lower organisms such as *C. elegans* and *Neurospora* (Bartel 2004; Baulcombe 2004; Matzke and Birchler 2005; Meister and Tuschl 2004). It involves the processing of double-stranded RNA (dsRNA) into small RNA (sRNA) species of 21 to 25 nucleotides (Hannon 2002). The dsRNA molecules can be generated by replicative intermediates of viral RNAs, inverted repeat homologous sequences, antisense or by aberrant RNAs, which become dsRNA by host RNA-dependent RNA polymerases (RdRPs or RDPs) (Ahlquist 2002; Dalmay et al. 2000; Waterhouse et al. 2001). These dsRNAs are cleaved by RNaseIII-like enzymes such as Dicer-like proteins (DCLs) (Bernstein et al. 2001) into small RNAs duplexes (sRNAs) of 21–25 nucleotides (nt) RNA fragments (Tang et al. 2003) which have 5’ phosphate and 2-nucleotide 3’ overhangs. These sRNAs can be divided into two major types, the small interfering RNAs (siRNAs) and the micro RNAs (miRNAs), based on their origin and formation. Subsequently, a guide strand (antisense to the target RNA) of a small RNA is incorporated into an RNA-induced silencing complex (RISC) which contains a member of the Argonaute (Ago) protein family. The RISC complex is responsible for the targeting and cleavage of sequence specific RNA and mediates translation repression (Elbashir et al. 2001; Tuschl et al. 1999; Zamore et al. 2000).

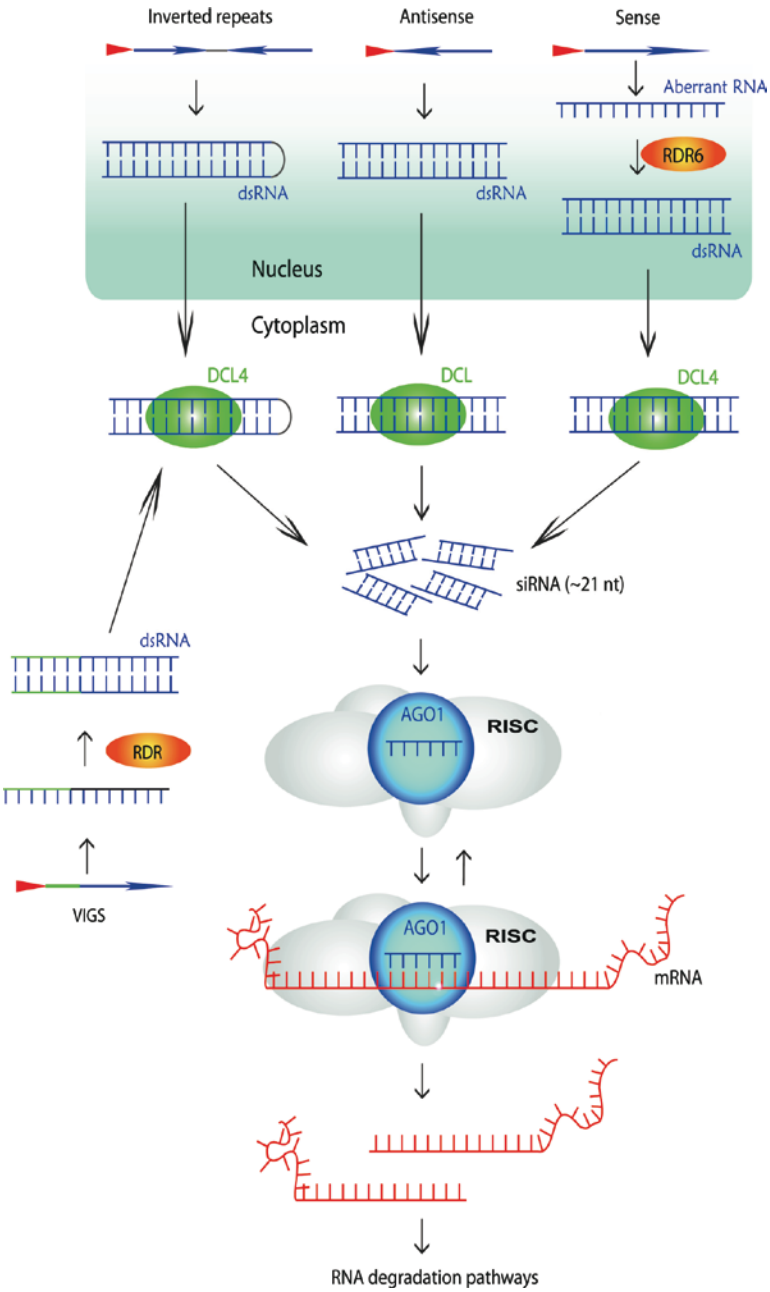


Fig. 27.1 Small RNA pathways involved in gene silencing. Long double stranded RNAs (dsRNAs) are first generated through the action of RNA dependent RNA polymerases. dsRNAs are then processed by Dicer-like enzymes (DCL) yielding short interfering RNAs. One strand of the siRNA then assembles into the RNA induced silencing complex (RISC) while the other strand is degraded. siRNA within the RISC then guides the cleavage of complementary mRNA performed by Argonaute 1 protein (Ago1). The resulting halves of mRNA are then degraded by other cellular enzymes

The number of DCL and AGO family members varies greatly among organisms. In *Arabidopsis*, at least 4 DCL and 10 AGO proteins have been identified (Brodersen and Voinnet 2006). Specific DCL and AGO proteins are proposed to have distinct roles within the various RNA-silencing pathways.

27.7 Plant Viruses and RNA Silencing

Plant viruses are known as strong inducers as well as targets of PTGS. In virus-induced silencing, the dsRNA either is a viral replication intermediate or is produced by a host-encoded RNA polymerase using a viral RNA template (Mourrain et al. 2000). The dsRNA generates small interfering RNAs (siRNAs), which leads to the degradation of viral RNA by RNA silencing mechanism (Hamilton and Baulcombe 1999; Ratcliff et al. 1999; Szittya et al. 2002).

To counteract RNA silencing, many plant viruses produce silencing-suppressor proteins that interfere with various steps of the silencing machinery (Baulcombe 2004; Ding et al. 2004; Li and Ding 2001; Roth et al. 2004; Silhavy and Burgyan 2004; Voinnet 2005). Silencing-suppressor proteins have been identified in various plant virus families such as HC-Pro of potyvirus (Anandalakshmi et al. 1998; Brigneti et al. 1998; Kasschau and Carrington 1998), p19 from tombusvirus (Lakatos et al. 2004; Silhavy et al. 2002), p21 from Beet yellows virus (Reed et al. 2003), p25 from potato virus X (PVX) (Voinnet et al. 2000) and p122 from *Tobacco Mosaic Virus* (Csorba et al. 2007). Several silencing-suppressor proteins bind to siRNAs and prevent the interaction of siRNAs with either the RDP enzymes or the RISC complexes (Csorba et al. 2007; Lakatos et al. 2004; Merai et al. 2006). Silencing-suppressor proteins can also inhibit the methylation of several miRNAs and suppress the RNA silencing (Chapman et al. 2004; Lozsa et al. 2008; Vogler et al. 2007; Yu et al. 2006). The better understanding of action of silencing suppressor proteins at molecular level can improve our basic knowledge about the intimate plant-virus interactions and also provided valuable tools to unravel the diversity, regulation, and evolution of RNA-silencing pathways.

27.8 Virus Induced Gene Silencing (VIGS)

Virus induced gene silencing has been developed as a technique that uses viral vectors carrying a fragment of a gene of interest to generate a dsRNA to suppress endogenous plant gene expression in order to study plant gene functions (Burch-Smith et al. 2004). In this technique, a plant gene sequence is inserted into viral genome and the viral RNA, produced by transcription *in vitro*, is inoculated into plant. The VIGS technique was improved by constructing a binary vector based on tobacco rattle virus that was modified to facilitate insertion of non-viral sequence and subsequent infection of plants. This vector induces very mild symptoms, infects large areas of adjacent cells and silences gene expression in growing points

(Ratcliff et al. 2001). *A. tumefaciens* containing the binary vector can then be easily introduced into plant tissues by a variety of techniques, including infiltration, direct inoculation of bacterial colonies, agrodrenching, or vacuum infiltration (Burch-Smith et al. 2004; Hileman et al. 2005; Lu et al. 2003; Ryu et al. 2004; Wang et al. 2006). VIGS has been demonstrated for a number of RNA and DNA viruses (Li et al. 2002; Ratcliff et al. 1999; Vance and Vaucheret 2001).

VIGS is a powerful tool due to its fast initiation of silencing in intact wild-type or transgenic plants (Robertson 2004). It can be used in the mature plants without plant transformation step. VIGS has been adopted for gene analysis both in dicotyledonous species, including Arabidopsis (Burch-Smith et al. 2006; Wang et al. 2006), potato (Brigneti et al. 2004), tomato (Liu et al. 2002), pepper (Chung et al. 2004), and in monocot plants such as barley (Hein et al. 2005; Holzberg et al. 2002), wheat (Scofield et al. 2005) and rice (Ding et al. 2007; Ding et al. 2006). The limitation of this approach resides in the viral host range, pattern of viral spreading, and symptoms caused by the infection. It is not universally applicable and requires optimization for each species.

27.9 SiRNA and DNA Methylation

DNA methylation in higher eukaryotes occurs on cytosine at the fifth position of the pyrimidine ring (5mC) (Colot and Rossignol 1999; Jost and Saluz 1993). In mammals, DNA methylation mainly occurs on the cytosine in a CpG context, in plants the cytosine can be methylated in the CpG (CG), CpNpG (CNG), and CpNpN (CNN) context, where the letters "p" signifies a phosphodiester bond that connects the C and G, and "N" represents any nucleotide other than guanine.

In *Arabidopsis*, at least three classes of methyltransferases (MT) are involved in methylation: MET1, Domains Rearranged Methylase (DRM) and CMT3 (Chromomethylase 3) (Finnegan and Kovac 2000). MET1 is a homolog of mammalian DNMT1 and has a maintenance function of CG methylation (Finnegan and Dennis 1993; Finnegan et al. 1996; Ronemus et al. 1996). It can also contribute to CG *de novo* methylation in the presence of RNA signals (Aufsatz et al. 2004). DRM shows similarity to mammalian DNMT3 (Cao et al. 2000). DRM2 is implicated in *de novo* DNA methylation. CMT3, unique to the plant kingdom, is involved in methylation at CNG motifs (Bartee et al. 2001). DRM2 along with MET1 and CMT3 are involved in maintaining methylation marks through DNA replication (Cao and Jacobsen 2002).

RNA-mediated silencing acts post-transcriptionally to direct translational repression of target RNAs, but can also be involved in nuclear processes leading to DNA methylation and/or heterochromatin formation (Wassenegger 2005). In plants, there are two functionally distinct siRNA species that likely arise from separate DICER-like activities (Hamilton et al. 2002; Tang et al. 2003). The 21–22 nt siRNAs trigger mostly PTGS and are believed to be involved in short distance signaling, while the 24–26 nt siRNAs trigger TGS, associated with DNA methylation and the systemic spread of silencing (Hamilton et al. 2002; Himber et al. 2003;

Kasschau et al. 2007; Xie et al. 2004). The siRNAs that direct methylation at endogenous DNA repeats in Arabidopsis are mainly produced by RNA-dependent RNA polymerase II (RDR2), RNA polymerase IV (Pol IV), RNA polymerase V, Dicer-like 3 (DCL3) and Argonaute4 (AGO4) (Haag et al. 2009; Matzke et al. 2007; Pontes et al. 2006; Wierzbicki et al. 2008).

RNA-directed DNA methylation (RdDM) was first discovered in transgenic tobacco plants that carried the potato spindle tuber viroid sequence as a transgene (Wassenegger et al. 1994). RdDM requires DRM2 and the SNF2-like protein, Defective in RNA Directed DNA Methylation 1 (DRD1) (Kanno et al. 2004). RdDM is also correlated with histone H3K9 methylation within the target sequences of endogenous repeat loci, and in the model transgenes that serve as targets of RdDM in arabidopsis (Huettel et al. 2006; Xie et al. 2004; Zilberman et al. 2004).

In rice, an examination of siRNA activity in DNA methylation and chromatin modifications using siRNA that targeted to one transgene and seven endogenous genes indicated that siRNAs induce *de novo* DNA methylation of target promoters, but do not induce transcriptional suppression (Okano et al. 2008). Analysis of epigenetic modifications induced by siRNA targeted to transcribed regions of endogenous rice genes showed that the de novo DNA methylation is gene-specific but it is insufficient to induce heterochromatin formation (Miki and Shimamoto 2008).

27.10 microRNA (MiRNA)

MicroRNAs are ~22 nt endogenous RNAs that can play important regulatory roles in animals and plants by cleavage or translational repression of mRNA (Bartel 2004; Jones-Rhoades and Bartel 2004). miRNA was discovered in early 1990s in the worm *C. elegans* by Victor Ambros and his colleagues. They were studying the role of *lin-4* and *lin-14* genes involved in developmental process. It was found that *lin-4*, which controls the expression of *lin-14*, did not code for a protein but instead produced a pair of short RNA molecules. Because these short RNAs had a part of their sequence similar to *lin-14*, it was postulated that the translational repression of *lin-14* is caused by sequence complementarity between lin-4 and unique repeats within the 3′ UTR of the *lin-14* mRNA (Lee et al. 1993). Second, miRNA, *let-7*, was discovered 7 years later in the same organism and it is also involved in developmental process (Reinhart et al. 2000). These two startling discoveries would probably have remained unknown had not several genomes been sequenced at that time. Using *in silico* approaches in combination with random cloning and sequencing it was soon discovered that thousands of miRNAs exist in species ranging from viruses to primates (Enright et al. 2003; Kiriakidou et al. 2004; Krek et al. 2005; Lewis et al. 2003, 2005; Rhoades et al. 2002). Today, a publicly accessible repository of identified miRNAs up to date is hosted by the Sanger institute (http://microrna.sanger.ac.uk/).

After a decade of extensive research, we are now beginning to understand the scope and diversity of these regulatory molecules. miRNAs are double-stranded

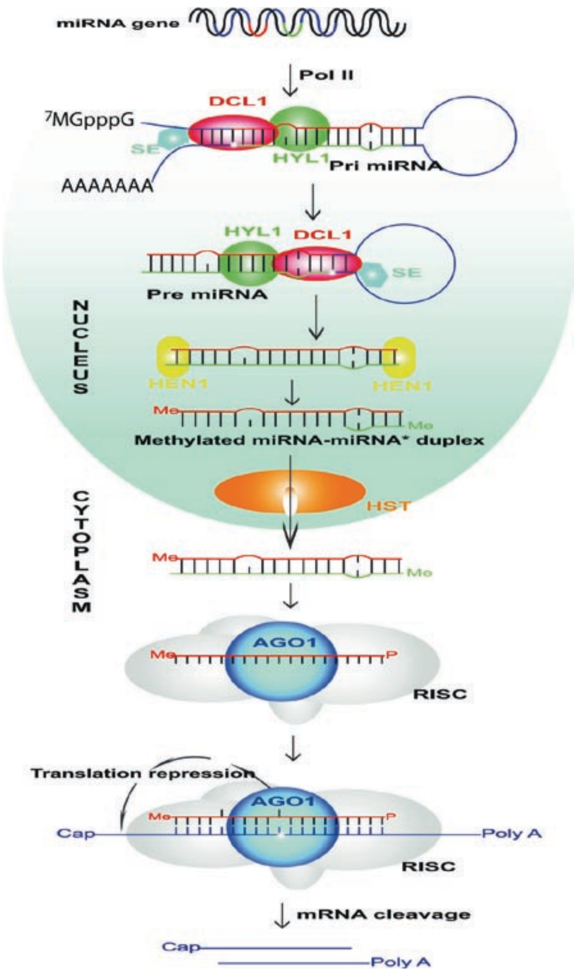


Fig. 27.2 miRNA biogenesis and action in plants. Micro RNAs are processed from specific DNA encoded precursors. These precursors form intramolecular hairpins containing imperfect base pairing. For detailed explanation see the text

RNAs of 21–25 nt that are derived from endogenously expressed transcripts with characteristic hairpin (hp) structures. The miRNA pathway (Fig. 27.2) begins with the transcription of a primary miRNA (pri-miRNA) from a miRNA gene. The pri-miRNA transcripts are generally 70–250 nt hp RNAs, that are longer than the miRNA stem loop itself (Xie et al. 2005). Subsequently, pri-miRNAs are processed in small, often perinucleolar bodies that are distinct from the Cajal bodies that serve as centers for the assembly of protein/siRNA silencing complexes in Arabidopsis (Song et al. 2007).

In the cytoplasm, the miRNA binds the RNA-Induced Silencing Complex (RISC) and aligns with the mRNA. RISC assembly is the core part of the RNAi

and miRNA pathways. Numbers of putative RISC proteins were identified by biochemical and genetical studies. The key protein of this complex is the Argonaute protein (Ago1). Argonautes were discovered a decade ago and originally thought to be involved in plant development (Bohnert et al. 1998; Moussian et al. 1998). Their involvement in miRNA and RNAi regulatory pathways later generated a great interest in their function. Among all three paralogous Argonaute groups (Hutvagner and Simard 2008), out of ten Ago proteins identified by homology searches, only Ago1 was found to be associated with the miRNA pathway in Arabidopsis (Vaucheret et al. 2004). The Ago1 protein is the active center of the RISC complex and it is apparently responsible for recognizing the guide strand in the miRNA/miRNA* duplex (Baumberger and Baulcombe 2005). The recruitment of miRNA into the RISC activates the Slicer activity of Ago1 and consequently initiates the transcript cleavage. In plants, this is the predominant way by which gene expression is repressed. In mammalian cells, due to the limited miRNA:mRNA base pairing, the predominant mode of action is translational inhibition (Mallory and Vaucheret 2006).

A variety of processes, including development, cell proliferation and death, are regulated by miRNAs and, in mammalian cells, have been recently linked to oncogenesis (Llave et al. 2002; Palatnik et al. 2003; Rhoades et al. 2002; Yekta et al. 2004). The identification and characterization of miRNAs is a rapidly growing area of research. Microarray-based expression analysis promises to be an ideal strategy for identifying candidate miRNAs that correlate with biological pathways and for generating molecular signatures of disease states. If a miRNA is found to be differentially expressed in a certain tissue or cell type, it may be hypothesized to play a regulatory role in specifying tissue or cell identity. Similarly, if a miRNA is expressed at a specific developmental stage, then it may regulate developmental timing (Du and Zamore 2005). miRNA expression can be profiled by the cloning and sequencing of miRNAs from specific tissues or cells, or at a specific developmental stage, or by microarray analysis (Barad et al. 2004; Baskerville and Bartel 2005). miRNA microarrays have been successfully used for the primary identification of new miRNA that are predicted by bioinformatics approaches, for dissection of differential expression of different miRNAs in the same cells and for the comparison of miRNA expression profiles from different tissues and cells (for review see (Yin et al. 2008). In plants however, due to the nearly perfect complementation of miRNAs to their target sequence, prediction through bioinformatics approaches is the most successful tool (Jones-Rhoades and Bartel 2004; Llave et al. 2002; Reinhart et al. 2002). In addition to bioinformatics, cloning and sequencing miRNAs is a widely used, if laborious, method for the discovery of new miRNAs. However, new emerging techniques, which avoid the need of cloning, may change this disadvantage in the near future. Massive parallel signature sequencing was recently applied to small RNA sequencing. The authors claimed that nearly all expressed miRNAs present in the sample were revealed by this technique in rice and Arabidopsis (Lu et al. 2005, 2008).

27.11 RNA Silencing for Crop Improvement

RNA-mediated silencing is proving to be a valuable tool for investigation of gene function. In Australia, Peter Waterhouse and his CSIRO group pioneered the use of RNAi technology to develop varieties of barley that are resistant to *barley yellow dwarf virus* (BYDV) (Wang et al. 2000). They developed several RNAi vectors by replacing the loop in hairpin RNA interference (hpRNAi) with an intron which the efficiency of gene silencing can be enhanced from about 50% to nearly 100% (Helliwell and Waterhouse 2003; Wesley et al. 2001). Kusaba and colleagues (Kusaba et al. 2003) applied RNAi to reduce the level of glutenin in rice and produced a LGC-1 (low glutenin content 1) rice variety. This low-protein rice is useful for patients with kidney disease whose protein intake is restricted. The trait was stable and was transmitted for a number of generations.

Recombinant DNA engineering is having remarkable success in overcoming plant pathogens. A spectacular example using gene silencing is the rescue of the Hawaiian papaya industry by conferring resistance to papaya ringspot virus (PRSV) (Fuchs and Gonsalves 2007). Another notable achievement is the bioengineered resistance of "NewLeaf Plus" potatoes to Potato leafroll virus, released by Monsanto (Lawson et al. 2001).

An important challenge for the future is the development of rice resistant to tungro disease. This disease, caused by the combined activities of a dsDNA virus (rice tungro bacilliform virus) and rice tungro spherical virus, a ssRNA virus, is responsible for annual losses approaching $US 10^9. Recently, Tyagi et al. (2008) transformed rice plants with DNA encoding ORF IV of RTBV in sense and anti-sense orientations. RNA blot analysis of two representative lines showed degradation of the transgene transcripts and accumulation of small molecular weight RNA. Following inoculation, in transgenic line (RTBV-O-Ds1) there was an initial rapid buildup of RTBV levels, comparable to that of untransformed controls, followed by a sharp reduction, resulting in a 50-fold decease in viral titer. In contrast, untransformed controls maintained a high viral titer for 40 days. In line RTBV-O-Ds2, RTBV DNA levels rose gradually to almost 60% of the control by 40 dpi. The Ds1 line showed tungro symptoms similar to the control lines, whereas line Ds2 showed extremely mild symptoms. These results provide preliminary evidence that RNAi approaches will ultimately be effective against tungro.

27.12 Artificial miRNA (AmiRNA) – an Emerging Approach of Great Promise

Schwab et al. (2006) described pioneering experiments using amiRNA in Arabidopsis. The natural target sequence for *A. thaliana* TPC transcription factor miR319a was replaced by a sequence of choice using ligation-mediated PCR to yield an amiRNA of choice. The procedure is straightforward and highly efficient

(Ossowski et al. 2008; Schwab et al. 2006). PCR primers can now be designed free of charge using an on-line tool available at http://wmd2.weigelworld.org.

The use of amiRNA shows great promise for agricultural applications. Niu et al. (2006) showed that transgenic Arabidopsis with separate amiRNA constructs targeting either TYMV or TuMV coat proteins were resistant to the cognate virus. Expression of a dimeric construct targeting each coat protein yielded plants resistant to both viruses, thereby demonstrating applicability of the amiRNA approach for engineering multiple virus resistance in crop plants.

amiRNA-induced gene silencing was shown to be more effective than the hpRNAi system (Qu et al. 2007). Original analysis of the efficacy of amiRNA for gene silencing (http://wmd2.weigelworld.org) suggested that ~75% of amiRNAs designed using the parameters given in (Schwab et al. 2006) were effective. Recently published studies however suggest that the success rate for amiRNA gene silencing is up to 90% (Alvarez et al. 2006; Choi et al. 2007; Mathieu et al. 2007; Ossowski et al. 2008), and even higher efficiencies can be envisaged by using two amiRNAs per target gene.

Scientists at the International Rice Research Institute (IRRI) have employed the amiRNA approach to target three different rice genes, *Phytoene desaturase* (*Pds*, Os03g08570), *Spotted leaf 11* (*Spl11*, Os12g38210), and *Elongated uppermost internode1* (*Eui1 /CYP714D1*, Os05g40384) *Pds, Spl11, and Eui1/CYP714D1* (Warthmann et al. 2008). amiRNA vectors derived from an endogenous rice miRNA precursor, osa-MIR528 were constructed and transformed into two rice varieties, Nipponbare (japonica) and IR64 (indica). In both varieties, the targeted genes in transgenic lines containing the amiRNA are down-regulated by amiRNA-guided cleavage of the transcripts, resulting in the expected mutant phenotypes as an albino phenotype (*pds*), spontaneous lesion formation in the absence of pathogens (*spl11*), and elongation of the uppermost internode at heading stage (*eui1*) respectively. The effects are highly specific to the target gene, the transgenes are stably inherited and they remain effective in the progeny.

This research showed that amiRNAs can efficiently trigger gene silencing in a monocot crop and can effectively modulate agronomically important traits in varieties used in modern breeding programs. The approach is suited for candidate gene validation, comparative functional genomics between different varieties, and for improvement of agronomic performance and nutritional value.

Acknowledgement Support to TCH from NSF (MCB 0420102; MCB 0346681) for aspects of the studies covered here is gratefully acknowledged.

References

Ahlquist P (2002) RNA-Dependent RNA polymerases, viruses, and RNA silencing. Science 296:1270–1273

Alvarez JP, Pekker I, Goldshmidt A, Blum E, Amsellem Z, Eshed Y (2006) Endogenous and synthetic microRNAs stimulate simultaneous, efficient, and localized regulation of multiple targets in diverse species. Plant Cell 18:1134–1151

Anandalakshmi R, Pruss GJ, Ge X, Marathe R, Mallory AC, Smith TH, Vance VB (1998) A viral suppressor of gene silencing in plants. Proc Natl Acad Sci U S A 95:13079–13084

Assaad FF, Tucker KL, Signer ER (1993) Epigenetic repeat-induced gene silencing (RIGS) in *Arabidopsis*. Plant Mol Biol 22:1067–1085

Aufsatz W, Mette M, Matzke A, Matzke M (2004) The role of MET1 in RNA-directed *de novo* and maintenance methylation of CG dinucleotides. Plant Mol Biol 54:793–804

Barad O, Meiri E, Avniel A, Aharonov R, Barzilai A, Bentwich I, Einav U, Gilad S, Hurban P, Karov Y, Lobenhofer EK, Sharon E, Shiboleth YM, Shtutman M, Bentwich Z, Einat P (2004) MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. Genome Res 14:2486–2494

Bartee L, Malagnac F, Bender J (2001) Arabidopsis *cmt3* chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. Genes Dev 15:1753–1758

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297

Baskerville S, Bartel DP (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. RNA 11:241–247

Baulcombe DC (1999) Viruses and gene silencing in plants. Arch Virol Suppl 15:189–201

Baulcombe DC (2004) RNA silencing in plants. Nature 431:356–363

Baulcombe DC (1996) Mechanisms of pathogen-derived resistance to viruses in transgenic plants. Plant Cell 8:1833–1844

Baulcombe DC, English JJ (1996) Ectopic pairing of homologous DNA and post-transcriptional gene silencing in transgenic plants. Curr Opin Biotechnol 7:173–180

Baumberger N, Baulcombe DC (2005) Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc Natl Acad Sci U S A 102:11928–11933

Bernstein E, Caudy AA, Hammond SM, Hannon GJ (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 409:363–366

Bestor TH (1998) The host defence function of genomic methylation patterns. Novartis Found Symp 214:187–195; discussion 195–189, 228–132

Bevan MW, Flavell RB, Chilton M-D (1983) A chimaeric antibiotic resistance gene as a selectable marker for plant cell transformation. Nature 304:184–187

Bohnert K, Camus I, Bellini C, Bouchez D, Caboche M, Benning C (1998) *AGO1* defines a novel locus of Arabidopsis controlling leaf development. EMBO J 17:170–180

Bollman KM, Aukerman MJ, Park MY, Hunter C, Berardini TZ, Poethig RS (2003) HASTY, the Arabidopsis ortholog of exportin 5/MSN5, regulates phase change and morphogenesis. Development 130:1493–1504

Brederode FT, Taschner PE, Posthumus E, Bol JF (1995) Replicase-mediated resistance to alfalfa mosaic virus. Virology 207:467–474

Brigneti G, Voinnet O, Li WX, Ji LH, Ding SW, Baulcombe DC (1998) Viral pathogenicity determinants are suppressors of transgene silencing in *Nicotiana benthamiana*. EMBO J 17:6739–6746

Brigneti G, Martin-Hernandez AM, Jin H, Chen J, Baulcombe DC, Baker B, Jones JD (2004) Virus-induced gene silencing in Solanum species. Plant J 39:264–272

Brodersen P, Voinnet O (2006) The diversity of RNA silencing pathways in plants. Trends Genet 22:268–280

Burch-Smith TM, Anderson JC, Martin GB, Dinesh-Kumar SP (2004) Applications and advantages of virus-induced gene silencing for gene function studies in plants. Plant J 39:734–746

Burch-Smith TM, Schiff M, Liu Y, Dinesh-Kumar SP (2006) Efficient virus-induced gene silencing in Arabidopsis. Plant Physiol 142:21–27

Cao J, Duan X, McEiroy D, Wu R (1992) Regeneration of herbicide resistant transgenic rice plants following microprojectile-mediated transformation of suspension culture cells. Plant Cell Rep 11:586–591

Cao X, Jacobsen SE (2002) Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc Natl Acad Sci U S A 99 Suppl. 4:16491–16498

Cao X, Springer NM, Muszynski MG, Phillips RL, Kaeppler S, Jacobsen SE (2000) Conserved plant genes with similarity to mammalian *de novo* DNA methyltransferases. Proc Natl Acad Sci U S A 97:4979–4984

Chapman EJ, Prokhnevsky AI, Gopinath K, Dolja VV, Carrington JC (2004) Viral RNA silencing suppressors inhibit the microRNA pathway at an intermediate step. Genes Dev 18:1179–1186

Chareonpornwattana S, Thara KV, Wang L, Datta SK, Panbangred W, Muthukrishnan S (1999) Inheritance, expression, and silencing of a chitinase transgene in rice. Theor Appl Genet 98:371–378

Cheng X, Sardana R, Kaplan H, Altosaar I (1998) *Agrobacterium*-transformed rice plants expressing synthetic cryIA(b) and cryIA(c) genes are highly toxic to striped stem borer and yellow stem borer. Proc Natl Acad Sci U S A 95:2767–2772

Choi K, Park C, Lee J, Oh M, Noh B, Lee I (2007) Arabidopsis homologs of components of the SWR1 complex regulate flowering and plant development. Development 134:1931–1941

Chung E, Seong E, Kim YC, Chung EJ, Oh SK, Lee S, Park JM, Joung YH, Choi D (2004) A method of high frequency virus-induced gene silencing in chili pepper (*Capsicum annuum* L. cv. Bukang). Mol Cells 17:377–380

Cogoni C, Macino G (1999) Posttranscriptional Gene Silencing in Neurospora by a RecQ DNA Helicase. Science 286:2342–2344

Colot V, Rossignol JL (1999) Eukaryotic DNA methylation as an evolutionary device. Bioessays 21:402–411

Csorba T, Bovi A, Dalmay T, Burgyan J (2007) The p122 subunit of Tobacco Mosaic Virus replicase is a potent silencing suppressor and compromises both small interfering RNA- and microRNA-mediated pathways. J Virol 81:11768–11780

Dalmay T, Hamilton A, Rudd S, Angell S, Baulcombe DC (2000) An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. Cell 101:543–553

Depicker A, Montagu MV (1997) Post-transcriptional gene silencing in plants. Curr Opin Cell Biol 9:373–382

Ding SW, Li H, Lu R, Li F, Li WX (2004) RNA silencing: a conserved antiviral immunity of plants and animals. Virus Res 102:109–115

Ding XS, Schneider WL, Chaluvadi SR, Mian MA, Nelson RS (2006) Characterization of a Brome mosaic virus strain and its use as a vector for gene silencing in monocotyledonous hosts. Mol Plant Microbe Interact 19:1229–1239

Ding XS, Rao CS, Nelson RS (2007) Analysis of gene function in rice through virus-induced gene silencing. Methods Mol Biol 354:145–160

Dorer DR, Henikoff S (1997) Transgene repeat arrays interact with distant heterochromatin and cause silencing in *cis* and *trans*. Genetics 147:1181–1190

Du T, Zamore PD (2005) microPrimer: the biogenesis and function of microRNA. Development 132:4645–4652

Elbashir SM, Martinez J, Patkaniowska A, Lendeckel W, Tuschl T (2001) Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. EMBO J 20:6877–6888

English JJ, Mueller E, Baulcombe DC (1996) Suppression of virus accumulation in transgenic plants exhibiting silencing of nuclear genes. Plant Cell 8:179–188

Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS (2003) MicroRNA targets in Drosophila. Genome Biol 5:R1

Finnegan EJ, Dennis ES (1993) Isolation and identification by sequence homology of a putative cytosine methyltransferase from *Arabidopsis thaliana*. Nucleic Acids Res 21:2383–2388

Finnegan EJ, Kovac KA (2000) Plant DNA methyltransferases. Plant Mol Biol 43:189–201

Finnegan EJ, Peacock WJ, Dennis ES (1996) Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. Proc Natl Acad Sci U S A 93:8449–8454

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391:806–811

Fraley RT, Rogers SG, Horsch RB, Sanders PR, Flick JS, Adams SP, Bittner ML, Brand LA, Fink CL, Fry JS, Galluppi GR, Goldberg SB, Hoffmann NL, Woo SC (1983) Expression of bacterial genes in plant cells. Proc Natl Acad Sci U S A 80:4803–4807

Fuchs M, Gonsalves D (2007) Safety of virus-resistant transgenic plants two decades after their introduction: lessons from realistic field risk assessment studies. Annu Rev Phytopathol 45:173–202

Garrick D, Fiering S, Martin DI, Whitelaw E (1998) Repeat-induced gene silencing in mammals. Nat Genet 18:56–59

Gaudet J, VanderElst I, Spence AM (1996) Post-transcriptional regulation of sex determination in *Caenorhabditis elegans*: widespread expression of the sex-determining gene fem-1 in both sexes. Mol Biol Cell 7:1107–1121

Goodwin J, Chapman K, Swaney S, Parks TD, Wernsman EA, Dougherty WG (1996) Genetic and biochemical dissection of transgenic RNA-mediated virus resistance. Plant Cell 8:95–105

Haag JR, Pontes O, Pikaard CS (2009) Metal A and Metal B sites of nuclear RNA polymerases Pol IV and Pol V are required for siRNA-dependent DNA methylation and gene silencing. PLoS ONE 4:e4110

Hamilton A, Voinnet O, Chappell L, Baulcombe D (2002) Two classes of short interfering RNA in RNA silencing. EMBO J 21:4671–4679

Hamilton AJ, Baulcombe DC (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 286:950–952

Hannon GJ (2002) RNA interference. Nature 418:244–251

Hein I, Barciszewska-Pacak M, Hrubikova K, Williamson S, Dinesen M, Soenderby IE, Sundar S, Jarmolowski A, Shirasu K, Lacomme C (2005) Virus-induced gene silencing-based functional characterization of genes associated with powdery mildew resistance in barley. Plant Physiol 138:2155–2164

Helliwell C, Waterhouse P (2003) Constructs and methods for high-throughput gene silencing in plants. Methods 30:289–295

Herrera-Estrella L, Depicker A, Van Montagu M, Schell J (1983) Expression of chimaeric genes transferred into plant cells using a Ti-plasmid-derived vector. Nature 303:209–213

Hiei Y, Ohta S, Komari T, Kumashiro T (1994) Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. Plant J 6:271–282

Hileman LC, Drea S, Martino G, Litt A, Irish VF (2005) Virus-induced gene silencing is an effective tool for assaying gene function in the basal eudicot species *Papaver somniferum* (opium poppy). Plant J 44:334–341

Himber C, Dunoyer P, Moissiard G, Ritzenthaler C, Voinnet O (2003) Transitivity-dependent and -independent cell-to-cell movement of RNA silencing. EMBO J 22:4523–4533

Holzberg S, Brosio P, Gross C, Pogue GP (2002) Barley stripe mosaic virus-induced gene silencing in a monocot plant. Plant J 30:315–327

Huettel B, Kanno T, Daxinger L, Aufsatz W, Matzke AJ, Matzke M (2006) Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. EMBO J 25:2828–2836

Hutvagner G, Simard MJ (2008) Argonaute proteins: key players in RNA silencing. Nat Rev Mol Cell Biol 9:22–32

Jacobs JJMR, Litière K, van Dijk V, van Eldik GJ, Van Montagu M, Cornelissen M (1997) Post-transcriptional β-1,3-glucanase gene silencing involves increased transcript turnover that is translation-independent. Plant J 12:885–893

Jones AR, Schedl T (1995) Mutations in gld-1, a female germ cell-specific tumor suppressor gene in *Caenorhabditis elegans*, affect a conserved domain also found in Src-associated protein Sam68. Genes Dev 9:1491–1504

Jones-Rhoades MW, Bartel DP (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Molecular Cell 14:787–799

Jost JP, Saluz HP (1993) Steroid hormone dependent changes in DNA methylation and its significance for the activation or silencing of specific genes. EXS 64:425–451

Kanno T, Mette MF, Kreil DP, Aufsatz W, Matzke M, Matzke AJ (2004) Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. Curr Biol 14:801–805

Kasschau KD, Carrington JC (1998) A counterdefensive strategy of plant viruses: suppression of posttranscriptional gene silencing. Cell 95:461–470

Kasschau KD, Fahlgren N, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Carrington JC (2007) Genome-wide profiling and analysis of Arabidopsis siRNAs. PLoS Biol 5:e57

Kiriakidou M, Nelson PT, Kouranov A, Fitziev P, Bouyioukos C, Mourelatos Z, Hatzigeorgiou A (2004) A combined computational-experimental approach predicts human microRNA targets. Genes Dev 18:1165–1178

Klein TM, Wolf ED, Wu R, Sanford JC (1987) High-velocity microprojectiles for delivering nucleic acids into living cells. Nature 327:70–73

Kohli A, Griffiths S, Palacios N, Twyman RM, Vain P, Laurie DA, Christou P (1999) Molecular characterization of transforming plasmid rearrangements in transgenic rice reveals a recombination hotspot in the CaMV 35S promoter and confirms the predominance of microhomology mediated recombination. Plant J 17:591–601

Kooter JM, Matzke MA, Meyer P (1999) Listening to the silent genes: transgene silencing, gene regulation and pathogen control. Trends Plant Sci 4:340–347

Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da PI, Gunsalus KC, Stoffel M, Rajewsky N (2005) Combinatorial microRNA target predictions. Nat Genet 37:495–500

Kumpatla SP, Hall TC (1998a) Longevity of 5-azacytidine-mediated gene expression and re-establishment of silencing in transgenic rice. Plant Mol Biol 38:1113–1122

Kumpatla SP, Hall TC (1998b) Recurrent onset of epigenetic silencing in rice harboring a multi-copy transgene. Plant J 14:129–135

Kumpatla SP, Hall TC (1999) Organizational complexity of a rice transgene locus susceptible to methylation-based silencing. IUBMB Life 48:459–467

Kumpatla SP, Chandrasekharan MB, Iyer LM, Li G, Hall TC (1998) Genome intruder scanning and modulation systems and transgene silencing. Trends Plant Science 3:97–104

Kusaba M, Miyahara K, Iida S, Fukuoka H, Takano T, Sassa H, Nishimura M, Nishio T (2003) Low glutelin content1: a dominant mutation that suppresses the glutelin multigene family *via* RNA silencing in rice. Plant Cell 15:1455–1467

Lakatos L, Szittya G, Silhavy D, Burgyan J (2004) Molecular mechanism of RNA silencing suppression mediated by p19 protein of tombusviruses. EMBO J 23:876–884

Lapidot M, Gafny R, Ding B, Wolf S, Lucas WJ, Beachy RN (1993) A dysfunctional movement protein of tobacco mosaic virus that partially modifies the plasmodesmata and limits virus spread in transgenic plants. Plant J 4:959–970

Lawson C, Weiss J, Thomas P, Kaniewski W (2001) NewLeaf Plus Russet Burbank potatoes: Replicase-mediated resistance to potato leafroll virus. Mol Breed 7:1–12

Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to *lin-14*. Cell 75:843–854

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN (2003) The nuclear RNase III Drosha initiates microRNA processing. Nature 425:415–419

Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB (2003) Prediction of mammalian microRNA targets. Cell 115:787–798

Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120:15–20

Li H, Li WX, Ding SW (2002) Induction and suppression of RNA silencing by an animal virus. Science 296:1319–1321

Li J, Yang Z, Yu B, Liu J, Chen X (2005) Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. Curr Biol 15:1501–1507

Li WX, Ding SW (2001) Viral suppressors of RNA silencing. Curr Opin Biotechnol 12:150–154

Li Z, Burow M, Murai N (1990) High frequency generation of fertile transgenic rice plants after PEG-mediated protoplast transformation. Plant Mol Biol Rep 8:276–291

Lindbo JA, Dougherty WG (1992) Pathogen-derived resistance to a potyvirus: immune and resistant phenotypes in transgenic tobacco expressing altered forms of a potyvirus coat protein nucleotide sequence. Mol Plant Microbe Interact 5:144–153

Lindbo JA, Silva-Rosales L, Proebsting WM, Dougherty WG (1993) Induction of a highly specific antiviral state in transgenic plants: Implications for regulation of gene expression and virus resistance. Plant Cell 5:1749–1759
Liu Y, Schiff M, Dinesh-Kumar SP (2002) Virus-induced gene silencing in tomato. Plant J 31:777–786
Llave C, Kasschau KD, Rector MA, Carrington JC (2002) Endogenous and silencing-associated small RNAs in plants. Plant Cell 14:1605–1619
Lomonossoff GP (1995) Pathogen-derived resistance to plant viruses. Annu Rev Phytopathol 33:323–343
Lozsa R, Csorba T, Lakatos L, Burgyan J (2008) Inhibition of 3' modification of small RNAs in virus-infected plants require spatial and temporal co-expression of small RNAs and viral silencing-suppressor proteins. Nucleic Acids Res 36:4099–4107
Lu C, Fedoroff N (2000) A mutation in the Arabidopsis HYL1 gene encoding a dsRNA binding protein affects responses to abscisic acid, auxin, and cytokinin. Plant Cell 12:2351–2366
Lu C, Tej SS, Luo S, Haudenschild CD, Meyers BC, Green PJ (2005) Elucidation of the small RNA component of the transcriptome. Science 309:1567–1569
Lu C, Jeong DH, Kulkarni K, Pillay M, Nobuta K, German R, Thatcher SR, Maher C, Zhang L, Ware D, Liu B, Cao X, Meyers BC, Green PJ (2008) Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). Proc Natl Acad Sci U S A 105:4951–4956
Lu R, Martin-Hernandez AM, Peart JR, Malcuit I, Baulcombe DC (2003) Virus-induced gene silencing in plants. Methods 30:296–303
Mallory AC, Vaucheret H (2006) Functions of microRNAs and related small RNAs in plants. Nat Genet 38 Suppl:S31-S36
Mathieu J, Warthmann N, Kuttner F, Schmid M (2007) Export of FT protein from phloem companion cells is sufficient for floral induction in Arabidopsis. Curr Biol 17:1055–1060
Matzke M, Kanno T, Huettel B, Daxinger L, Matzke AJ (2007) Targets of RNA-directed DNA methylation. Curr Opin Plant Biol 10:512–519
Matzke MA, Birchler JA (2005) RNAi-mediated pathways in the nucleus. Nat Rev Genet 6:24–35
Matzke MA, Matzke A (1995) How and why do plants inactivate homologous (trans)genes? Plant Physiol 107:679–685
Matzke MA, Matzke AJ (1998) Epigenetic silencing of plant transgenes as a consequence of diverse cellular defence responses. Cell Mol Life Sci 54:94–103
Meister G, Tuschl T (2004) Mechanisms of gene silencing by double-stranded RNA. Nature 431:343–349
Merai Z, Kerenyi Z, Kertesz S, Magna M, Lakatos L, Silhavy D (2006) Double-stranded RNA binding may be a general plant RNA viral strategy to suppress RNA silencing. J Virol 80:5747–5756
Metzlaff M, O'Dell M, Cluster PD, Flavell RB (1997) RNA-mediated RNA degradation and chalcone synthase A silencing in petunia. Cell 88:845–854
Miki D, Shimamoto K (2008) *De novo* DNA methylation induced by siRNA targeted to endogenous transcribed sequences is gene-specific and *OsMet1*-independent in rice. Plant J 56:539–549
Montgomery MK, Fire A (1998) Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression. Trends Genet 14:255–258
Morino K, Olsen OA, Shimamoto K (1999) Silencing of an aleurone-specific gene in transgenic rice is caused by a rearranged transgene. Plant J 17:275–285
Mourrain P, Beclin C, Elmayan T, Feuerbach F, Godon C, Morel JB, Jouette D, Lacombe AM, Nikic S, Picault N, Remoue K, Sanial M, Vo TA, Vaucheret H (2000) Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. Cell 101:533–542
Moussian B, Schoof H, Haecker A, Jurgens G, Laux T (1998) Role of the ZWILLE gene in the regulation of central shoot meristem cell fate during Arabidopsis embryogenesis. EMBO J 17:1799–1809
Murai N, Kemp JD, Sutton DW, Murray MG, Slightom JL, Merlo DJ, Reichert NA, Sengupta-Gopalan C, Stock CA, Barker RF, Hall TC (1983) Phaseolin gene from bean is expressed after transfer to sunflower via tumor-inducing plasmid vectors. Science 222:476–482

Napoli C, Lemieux C, Jorgensen R (1990) Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in *trans*. Plant Cell 2:279–289

Niu QW, Lin SS, Reyes JL, Chen KC, Wu HW, Yeh SD, Chua NH (2006) Expression of artificial microRNAs in transgenic *Arabidopsis thaliana* confers virus resistance. Nat Biotechnol 24:1420–1428

Okano Y, Miki D, Shimamoto K (2008) Small interfering RNA (siRNA) targeting of endogenous promoters induces DNA methylation, but not necessarily gene silencing, in rice. Plant J 53:65–77

Ossowski S, Schwab R, Weigel D (2008) Gene silencing in plants using artificial microRNAs and other small RNAs. Plant J 53:674–690

Pal-Bhadra M, Bhadra U, Birchler JA (1997) Cosuppression in *Drosophila*: gene silencing of Alcohol dehydrogenase by white-Adh transgenes is Polycomb dependent. Cell 90:479–490

Pal-Bhadra M, Bhadra U, Birchler JA (1999) Cosuppression of nonhomologous transgenes in *Drosophila* involves mutually related endogenous sequences. Cell 99:35–46

Palatnik JF, Allen E, Wu X, Schommer C, Schwab R, Carrington JC, Weigel D (2003) Control of leaf morphogenesis by microRNAs. Nature 425:257–263

Papp I, Mette MF, Aufsatz W, Daxinger L, Schauer SE, Ray A, van der WJ, Matzke M, Matzke AJ (2003) Evidence for nuclear processing of plant micro RNA and short interfering RNA precursors. Plant Physiol 132:1382–1390

Park MY, Wu G, Gonzalez-Sulser A, Vaucheret H, Poethig RS (2005) Nuclear processing and export of microRNAs in Arabidopsis. Proc Natl Acad Sci U S A 102:3691–3696

Pinto YM, Kok RA, Baulcombe DC (1999) Resistance to rice yellow mottle virus (RYMV) in cultivated African rice varieties containing RYMV transgenes. Nat Biotechnol 17:702–707

Pontes O, Li CF, Nunes PC, Haag J, Ream T, Vitins A, Jacobsen SE, Pikaard CS (2006) The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. Cell 126:79–92

Powell-Abel P, Nelson R, De B, Hoffmann N, Rogers S, Fraley R, Beachy R (1986) Delay of disease development in transgenic plants that express the tobacco mosaic virus coat protein gene. Science 232:738–743

Prins M, Goldbach R (1996) RNA-mediated virus resistance in transgenic plants. Arch Virol 141:2259–2276

Qu J, Ye J, Fang R (2007) Artificial microRNA-mediated virus resistance in plants. J Virol 81:6690–6699

Ratcliff FG, MacFarlane SA, Baulcombe DC (1999) Gene silencing without DNA:RNA. rna-mediated cross-protection between viruses. Plant Cell 11:1207–1216

Ratcliff F, Martin-Hernandez AM, Baulcombe DC (2001) Technical advance. Tobacco rattle virus as a vector for analysis of gene function by silencing. Plant J 25:237–245

Reed JC, Kasschau KD, Prokhnevsky AI, Gopinath K, Pogue GP, Carrington JC, Dolja VV (2003) Suppressor of RNA silencing encoded by Beet yellows virus. Virology 306:203–209

Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. Nature 403:901–906

Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP (2002) MicroRNAs in plants. Genes Dev 16:1616–1626

Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, Bartel DP (2002) Prediction of plant microRNA targets. Cell 110:513–520

Robertson D (2004) VIGS vectors for gene silencing: many targets, many tools. Annu Rev Plant Biol 55:495–519

Ronemus MJ, Galbiati M, Ticknor C, Chen J, Dellaporta SL (1996) Demethylation-induced developmental pleiotropy in Arabidopsis. Science 273:654–657

Rossignol JL, Faugeron G (1994) Gene inactivation triggered by recognition between DNA repeats. Experientia 50:307–317

Roth BM, Pruss GJ, Vance VB (2004) Plant viral suppressors of RNA silencing. Virus Res 102:97–108

Ryu CM, Anand A, Kang L, Mysore KS (2004) Agrodrench: a novel and effective agroinoculation method for virus-induced gene silencing in roots and diverse Solanaceous species. Plant J 40:322–331

Sanford JC, Johnston SA (1985) The concept of parasite-derived resistance—Deriving resistance genes from the parasite's own genome. J Theor Biol 113:395–405

Schwab R, Ossowski S, Riester M, Warthmann N, Weigel D (2006) Highly specific gene silencing by artificial microRNAs in Arabidopsis. Plant Cell 18:1121–1133

Scofield SR, Huang L, Brandt AS, Gill BS (2005) Development of a virus-induced gene-silencing system for hexaploid wheat and its use in functional analysis of the *Lr21*-mediated leaf rust resistance pathway. Plant Physiol 138:2165–2173

Silhavy D, Burgyan J (2004) Effects and side-effects of viral RNA silencing suppressors on short RNAs. Trends Plant Sci 9:76–83

Silhavy D, Molnar A, Lucioli A, Szittya G, Hornyik C, Tavazza M, Burgyan J (2002) A viral protein suppresses RNA silencing and binds silencing-generated, 21- to 25-nucleotide double-stranded RNAs. EMBO J 21:3070–3080

Song L, H, M-H, Lesicka J, Fedoroff N (2007) *Arabidopsis* primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. Proc Natl Acad Sci U S A 104:5437–5442

Stam M, Viterbo A, Mol JN, Kooter JM (1998) Position-dependent methylation and transcriptional silencing of transgenes in inverted T-DNA repeats: implications for posttranscriptional silencing of homologous host genes in plants. Mol Cell Biol 18:6165–6177

Szittya G, Molnar A, Silhavy D, Hornyik C, Burgyán J (2002) Short defective interfering RNAs of tombusviruses are not targeted but trigger post-transcriptional gene silencing against their helper virus. Plant Cell 14:359–372

Tang G, Reinhart BJ, Bartel DP, Zamore PD (2003) A biochemical framework for RNA silencing in plants. Genes Dev 17:49–63

Tanzer MM, Thompson WF, Law MD, Wernsman EA, Uknes S (1997) Characterization of post-transcriptionally suppressed transgene expression that confers resistance to tobacco etch virus infection in tobacco. Plant Cell 9:1411–1423

Tuschl T, Zamore PD, Lehmann R, Bartel DP, Sharp PA (1999) Targeted mRNA degradation by double-stranded RNA *in vitro*. Genes Dev 13:3191–3197

Tyagi H, Rajasubramaniam S, Rajam M, Dasgupta I (2008) RNA-interference in rice against Rice tungro bacilliform virus results in its decreased accumulation in inoculated rice plants. Transgenic Res 17:897–904

Van Blokland R, Van der Geest N, Mol JNM, Kooter JM (1994) Transgene-mediated suppression of chalcone synthase expression in *Petunia hybrida* results from an increase in RNA turnover. Plant J 6:861–877

van der Krol AR, Mur LA, Beld M, Mol J, Stuitje AR (1990) Flavonoid genes in petunia: Addition of a limited number of gene copies may lead to a suppression of gene expression. Plant Cell 2:291–299

Vance V, Vaucheret H (2001) RNA Silencing in plants – defense and counterdefense. Science 292:2277–2280

Vaucheret H (2006) Post-transcriptional small RNA pathways in plants: mechanisms and regulations. Genes Dev 20:759–771

Vaucheret H, Fagard M (2001) Transcriptional gene silencing in plants: targets, inducers and regulators. Trends Genet 17:28–35

Vaucheret H, Beclin C, Elmayan T, Feuerbach F, Godon C, Morel JB, Mourrain P, Palauqui JC, Vernhettes S (1998) Transgene-induced gene silencing in plants. Plant J 16:651–659

Vaucheret H, Vazquez F, Crete P, Bartel DP (2004) The action of *ARGONAUTE1* in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. Genes Dev 18:1187–1197

Vazquez F, Vaucheret H, Rajagopalan R, Lepers C, Gasciolli V, Mallory AC, Hilbert JL, Bartel DP, Crete P (2004) Endogenous *trans*-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. Mol Cell 16:69–79

Vogler H, Akbergenov R, Shivaprasad PV, Dang V, Fasler M, Kwon MO, Zhanybekova S, Hohn T, Heinlein M (2007) Modification of small RNAs associated with suppression of RNA silencing by tobamovirus replicase protein. J Virol 81:10379–10388

Voinnet O (2005) Induction and suppression of RNA silencing: insights from viral infections. Nat Rev Genet 6:206–220

Voinnet O, Lederer C, Baulcombe DC (2000) A viral movement protein prevents spread of the gene silencing signal in *Nicotiana benthamiana*. Cell 103:157–167
Wang C, Cai X, Wang X, Zheng Z (2006) Optimisation of tobacco rattle virus-induced gene silencing in Arabidopsis. Funct Plant Biol 33:347–355
Wang M-B, Abbott DC, Waterhouse PM (2000) A single copy of a virus-derived transgene encoding hairpin RNA gives immunity to barley yellow dwarf virus. Mol Plant Pathol 1:347–356
Warthmann N, Chen H, Ossowski S, Weigel D, Herve P (2008) Highly specific gene silencing by artificial miRNAs in rice. PLoSONE 3:e1829
Wassenegger M (2005) The role of the RNAi machinery in heterochromatin formation. Cell 122:13–16
Wassenegger M, Heimes S, Riedel L, Sanger HL (1994) RNA-directed de novo methylation of genomic sequences in plants. Cell 76:567–576
Waterhouse PM, Wang MB, Finnegan EJ (2001) Role of short RNAs in gene silencing. Trends Plant Sci 6:297–301
Wesley SV, Helliwell CA, Smith NA, Wang M-B, Rouse DT, Liu Q, Gooding PS, Singh SP, Abbott D, Stoutjesdijk PA, Robinson SP, Gleave AP, Green AG, Waterhouse PM (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. Plant J 27:581–590
Wierzbicki AT, Haag JR, Pikaard CS (2008) Noncoding transcription by RNA polymerase Pol IVb/ Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell 135:635–648
Xie Z, Romano DM, Kovacs DM, Tanzi RE (2004) Effects of RNA interference-mediated silencing of gamma-secretase complex components on cell sensitivity to caspase-3 activation. J Biol Chem 279:34130–34137
Xie Z, Allen E, Fahlgren N, Calamar A, Givan SA, Carrington JC (2005) Expression of Arabidopsis miRNA genes. Plant Physiol 138:2145–2154
Yang G, Lee YH, Jiang Y, Kumpatla SP, Hall TC (2005) Organization, not duplication, triggers silencing in a complex transgene locus in rice. Plant Mol Biol 58:351–366
Yekta S, Shih IH, Bartel DP (2004) MicroRNA-directed cleavage of HOXB8 mRNA. Science 304:594–596
Yi R, Doehle BP, Qin Y, Macara IG, Cullen BR (2005) Overexpression of exportin 5 enhances RNA interference mediated by short hairpin RNAs and microRNAs. RNA 11:220–226
Yi R, Qin Y, Macara IG, Cullen BR (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev 17:3011–3016
Yin JQ, Zhao RC, Morris KV (2008) Profiling microRNA expression with microarrays. Trends Biotechnol 26:70–76
Yu B, Yang Z, Li J, Minakhina S, Yang M, Padgett RW, Steward R, Chen X (2005) Methylation as a crucial step in plant microRNA biogenesis. Science 307:932–935
Yu B, Chapman EJ, Yang Z, Carrington JC, Chen X (2006) Transgenically expressed viral RNA silencing suppressors interfere with microRNA methylation in Arabidopsis. FEBS Lett 580:3117–3120
Zamore PD, Tuschl T, Sharp PA, Bartel DP (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell 101:25–33
Zhang HM, Yang H, Rech EL, Golds TJ, Davis AS, Mulligan BJ, Cocking EC, Davey MR (1988) Transgenic rice plants produced by electroporation-mediated plasmid uptake into protoplasts. Plant Cell Rep 7:379–384
Zhou X, Carranco R, Vitha S, Hall TC (2005) The dark side of green fluorescent protein. New Phytol 168:313–322
Zilberman D, Cao X, Johansen LK, Xie Z, Carrington JC, Jacobsen SE (2004) Role of Arabidopsis *ARGONAUTE4* in RNA-directed DNA methylation triggered by inverted repeats. Curr Biol 14:1214–1220

Chapter 28
Plant RNAi and Crop Improvement

Masayuki Isshiki and Hiroaki Kodama

Abstract RNA interference (RNAi) is a sequence-specific gene silencing technology that is mediated by small interfering RNAs (siRNAs) as a guide of RNA cleavage, translational inhibition and DNA methylation. Most plant RNAi vectors produce hairpin RNAs that are immediately processed into double-stranded siRNAs. RNAi technology has been applied in the genetic engineering of important plant metabolites including starches, oils and storage proteins. RNAi is also useful for preventing plants from viral infection. Since siRNAs (or their precursor, double-stranded RNAs) can be transferred from plants to plant-feeding pests, RNAi has been used to engineer plants resistant to nematodes and insects. In this chapter, several RNAi applications concerning crop improvement are introduced.

28.1 Introduction

The revolution in sequencing technology in the last several years has enabled us to obtain sequence information with low costs and extremely high throughput. In fact, direct comparison of whole genome sequences among 100 human genomes is being planned by National Institutes of Health (Pop and Salzberg 2007). Similarly, such large-scale sequencing technology has been applied to the sequencing of expressed sequencing tags (ESTs), for example, ESTs of *Medicago truncatula* (Cheung et al. 2006). As sequence production is increasing, the demand of functional identification of unknown genes is also increasing. RNA interference (RNAi) can ectopically reduce the expression of target genes (Fire et al. 1998) and has been applied at a genome-wide scale for gene discovery in *Caenorhabditis elegans* (Sönnichsen

M. Isshiki
Kihara Institute for Biological Research, Yokohama City University, 641-12 Maioka-cho, Yokohama, 244-0813, Japan

H. Kodama (✉)
Graduate School of Horticulture, Chiba University, 648 Matsudo, Chiba 271-8501, Japan
e-mail: kodama@faculty.chiba-u.jp

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_28,

et al. 2005) and mammalian cells (e.g., Whitehurst et al. 2007). A large-scale resource consisting of the *Arabidopsis* RNAi transgenic lines has been prepared by the AGRIKOLA project (Arabidopsis Genomic RNAi Knock-out Line Analysis) (www.agrikola.org; Hilson et al. 2004). A resource of 1,500 RNAi plasmids by using a *M. truncatula* EST library is being prepared (www.medicago.org/rnai/). An *Arabidopsis* RNAi library (named amiRNA; artifical miRNA) provided by Drs. Hanon, Martienssen, McCombie and their colleagues has been released on the market, in which each of the 22,000 genes is targeted by three different amiRNAs (www.openbiosystems.com/RNAi/Arabidopsisthaliana amiRNA/). amiRNA is a newly developed RNAi vector system in which only one small RNA molecule per amiRNA vector is produced. The characteristics and application of amiRNA have been recently reviewed (Ossowski et al. 2008). The approaches mentioned above will serve as a powerful tool for functional genomics. In addition, RNAi is also used in the improvement of crop properties by reducing the expression of undesirable genes.

RNAi is a conserved inhibitory mechanism seen in most eukaryotic organisms, where 21~25-nucleotide (nt) long, small RNA molecules (so-called small interfering RNA, siRNA) play a key role in identification of target genes (Hamilton and Baulcombe 1999). RNAi is heritable in a dominant manner, and has been used for crop breeding. As most researchers expect, the RNAi technology has been applied for knockdown of the expression of undesirable genes, and for enhancement of viral resistance. Another unique application is emerging, namely knockdown of gene function of the plant pests. When siRNAs (or their precursor, double-stranded RNAs, dsRNAs) generated in plants are taken in by plant-feeding pests, these siRNAs (or dsRNAs) can spread from the intestine to the somatic tissues in the worms, where siRNAs interfere with the gene function in situ. This strategy is based on a finding that the interfering activity of siRNAs can be transported across cell boundaries in worms (Tabara et al. 1998). In this chapter, we focus on these applications of RNAi technology to crop breeding, including metabolic engineering, viral resistance and resistance against the plant-feeding pests.

28.2 Outline of the RNAi Pathway

In the RNAi pathway, siRNAs interfere with the expression of target genes in several different ways, namely mRNA cleavage (Song et al. 2004), translational inhibition (Broderson et al. 2008) and DNA methylation-mediated transcriptional inactivation (RNA-directed DNA methylation, RdDM; Matzke et al. 2007) (Fig. 28.1). siRNA molecules are generated from dsRNAs by an RNaseIII-like enzyme, Dicer (Bernstein et al. 2001). In plants, this enzyme species is called DICER-LIKE protein (DCL) (Baulcombe 2004). Thus, plant RNAi has been carried out by expression of dsRNAs in most cases. Plant RNAi vectors transcribe an inverted repeat that consists of the sense and antisense sequences corresponding to the target sequences (Chuang and Meyerowitz 2000). Since replication of inverted repeat sequences without a spacer

is unstable in *Escherichia coli*, the inverted repeat region is usually separated by a spacer fragment. After transcription, inverted repeat sequences form a dsRNA structure, the so-called stem, and a spacer fragment forms a loop of a hairpin RNA. The stem portion is processed into siRNAs by Dicer. In *Arabidopsis thaliana*, four *DCL* genes (*DCL1* to *DCL4*) have been identified. Dunoyer et al. (2007) found that DCL1 facilitates production of 21- and 24-nt siRNAs via excision of the stem-loop portion of a hairpin RNA. Then 21-nt siRNAs are produced through the DCL4-dependent pathway (Dunoyer et al. 2007) and 24-nt siRNAs are processed via the DCL3-dependent pathway (Xie et al. 2004). In the absence of DCL4, the stem of a hairpin

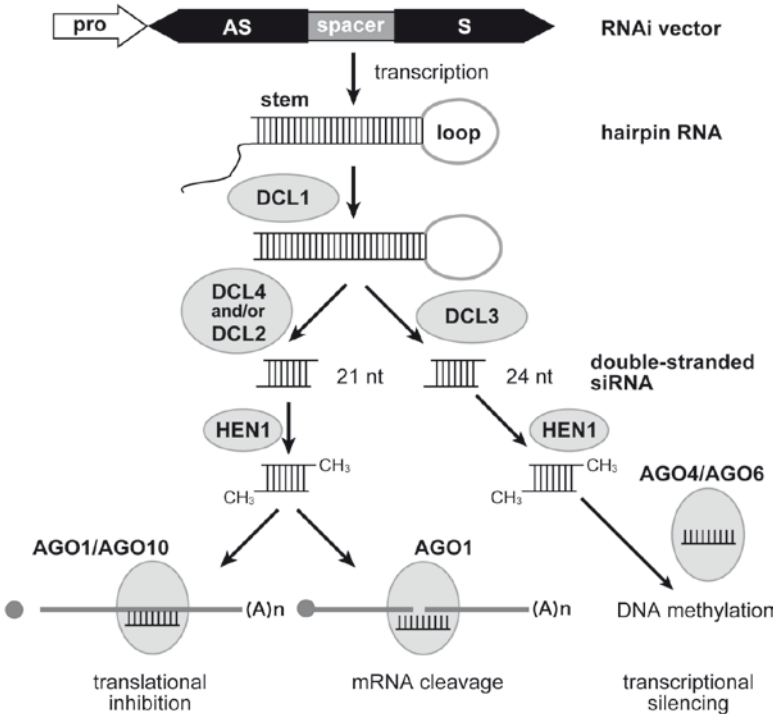


Fig. 28.1 Silencing of gene expression by RNAi pathway. An RNAi construct is comprised of a promoter and an inverted repeat interrupted by a spacer fragment. After transcription, a nascent transcript should have additional sequences originated from the terminator sequences and it is polyadenylated. DCL1 is considered to process this nascent transcript, and a 'mature' hairpin RNA is produced. The 21-nt long double-stranded siRNAs are produced by DCL4/DCL2 (see also text), and 24-nt long double-stranded siRNAs are generated by the action of DCL3. These siRNAs are then methylated at their 3′ termini by HEN1 (Yu et al. 2005), by which siRNAs are protected from degradation. A single-strand siRNA incorporated into AGO1 guides the mRNA cleavage. AGO1-siRNA and/or AGO10-siRNA complexes are considered to guide the translational inhibition, and a cap structure (indicated as a circle in this figure) is then removed. The siRNA-AGO4 complex has been shown to be necessary for sequence-specific DNA methylation, but its detailed mechanisms are mostly unknown. The siRNA-directed methylation of promoter sequences are followed by heterochromatinization and then transcription is inhibited

RNA is processed into 22-nt siRNAs by DCL2 (Dunoyer et al. 2007). When hairpin RNAs are transcribed under the control of a strong promoter such as the Cauliflower mosaic virus (CaMV) 35S promoter, both DCL2 and DCL4 are involved in the synthesis of 21-nt siRNAs from abundant hairpin RNAs (Fusaro et al. 2006; Mlotshwa et al. 2008). The products of DCL enzymes are double-stranded siRNA molecules, and one siRNA strand assembles into the ARGONAUTE (AGO) proteins. Ten distinct *AGO* paralogs have been identified in *Arabidopsis* plants. AGO1- and AGO4-associated small RNAs were predominantly 21- and 24-nt in length, respectively (Mi et al. 2008). AGO1 is a main slicer, an enzyme responsible for mRNA cleavage in the RNAi pathway (Baumberger and Baulcombe 2005), and AGO4 and AGO6 guide cytosine methylation in the genome (Zilberman et al. 2004; Zheng et al. 2007). Unlike animal cells, most plant siRNAs had been considered to guide exclusively mRNA cleavage but not translational inhibition. Recently, Brodersen et al. (2008) showed that siRNAs in AGO1 and/or AGO10 partially guide the repression of translation even though these siRNAs showed perfect complementarity. In most cases, it is difficult for us to detect translational repression because siRNA-mediated mRNA cleavage is dominant.

28.3 Plant RNAi Vectors

The construction of plant RNAi vectors is time-consuming and laborious work because sense and antisense target sequences and a spacer should be incorporated into a single vector. As a rapid and easy method for construction of RNAi vectors, Wesley et al. (2001) made a pHELLSGATE vector system that allows us to construct plant RNAi vectors by using Gateway technology (Invitrogen). In this system, PCR fragments with sense and antisense orientations are amplified by using primers having the recombinase recognition sites, *att*B1 and *att*B2 sequences. These PCR fragments are cloned into a plasmid containing *att*P1 and *att*P2 sites by a BP clonase. pHELLSGATE vector consists of a CaMV 35S promoter and a catalase intron as a spacer. pHELLSGATE can be directly used for *Agrobacterium*-mediated transformation of plants and facilitates high-throughput application of RNAi studies. Miki and Shimamoto (2004) developed another high-throughput RNAi vector system, the pANDA vector, in which a maize ubiquitin promoter is used for high expression of an RNAi cassette and a DNA fragment originated from the β-glucuronidase (*GUS*) gene was used as a spacer. To easily clone target DNA fragments, the Gateway technology is used in the pANDA vector system. A target PCR fragment containing CACC sequences at the 5′ end of forward primer is cloned into the pENTR/D-TOPO vector (Invitrogen), resulting that these PCR fragments are flanked with two recombination sites, *att*L1 and *att*L2. A LR clonase recombines this PCR fragment into two recombination sites (*att*R1 and *att*R2) of pANDA vector in opposite directions. pANDA vector is also ready for use in *Agrobacterium*-mediated transformation of monocots. The RNAi vectors developed by using Gateway technology are summarized in Table 28.1.

Table 28.1 RNAi vectors using Gateway technology

Vector	Promoter	Spacer	Selection marker	Transformation	References
pHELLSGATE	CaMV 35S	Pdk intron	NPTII	Agrobacterium infection	Wesley et al. (2001)
pHELLSGATE12	CaMV 35S	Pdk intron and Cat intron	NPTII	Agrobacterium infection	Helliwell and Waterhouse (2003)
pAGRIKOLA	CaMV 35S	Pdk intron and Cat intron	Bar	Agrobacterium infection	Hilson et al. (2004)
P*7GWIWG2	CaMV 35S	Intron	NPTII, HPT, Bar	Agrobacterium infection	Karimi et al. (2002)
pANDA	Maize Ubiquitin	GUS	NPTII, HPT	Agrobacterium infection	Miki and Shimamoto (2004)
pANDA-mini	Maize Ubiquitin	GUS	No marker	particle bombardment	Miki and Shimamoto (2004)
pIPKTA30N	CaMV 35S	RGA2 intron	No marker	particle bombardment	Douchkov et al. (2005)
pOpOff1	DEX-inducible promoter	Pdk intron and Cat intron	HPT	Agrobacterium infection	Wielopolska et al. (2005)
pIPKb007-pIPKb0010	Maize Ubiquitin, Rice Actin Enhanced CaMV 35S, Wheat Gst A1	RGA2 intron	HPT	Agrobacterium infection	Himmelbach et al. (2007)

Bar, bialaphos acetyltransferase; Cat, catalase; DEX, dexamethasone; GstA1, glutathione S-transferase A1; HPT, hygromycin phosphotransferase; NPTII, neomycin phosphotransferase II; Pdk, pyruvate orthophosphate dikinase; RGA2; a gene encoding a putative resistant protein.

The sequence specificity in RNAi is determined by hybridization of siRNAs to the corresponding target mRNAs. Knockdown of a single gene among the several paralog genes or simultaneous silencing of multiple genes have been demonstrated, in which the choice of target sequences is important. In addition, several inducible RNAi vectors have been developed. The strategies using RNAi vectors are summarized by Hirai and Kodama (2008).

28.4 RNAi-Mediated Metabolic Engineering

Plants are a source of numerous diet products and also important raw materials for industry, including papers, building materials and pharmaceuticals. Genetic approaches enhance our ability to improve crop traits. RNAi technology is now widely used and representative examples of its application to plant molecular breeding are shown in Table 28.2. Here, we discuss several examples in which RNAi was used as a tool of genetic manipulation of primary metabolites, including starch and oil, level control of the target proteins and modulation of secondary metabolisms in crop plants.

28.4.1 Starch Metabolism

Starch consists of two types of a glucan polymer that are called amylose and amylopectin. Amylose is a linear α-1,4 glucan, and is synthesized in amyloplast by granule-bound starch synthase (GBSS). Amylopectin, the main constituent of starch, is a highly branched glucan in which the glucose unit is joined by α-1,6 linkage. The starch-branching enzyme (SBE) is responsible for formation of α-1,6 linkage in amylopectin. Starch is a source of energy in diets and has been used as a renewable raw material (Jobling 2004). Physicochemical properties of starch can be altered through chemical and enzyme modifications, and such altered properties of starches are followed by development of unique applications. Thus, designing starch by genetic engineering of starch-synthesizing enzymes has been an important breeding subject.

28.4.1.1 Amylose-Free Starch

Amylose-free (waxy) starch has been produced by a corn *waxy* mutant. Corn waxy starch easily gelatinizes and yields clear pastes without gel formation. Heilersig et al. (2006) reported an inhibition of potato (*Solanum tuberosum* L.) granule-bound starch synthase I (GBSSI) activity by introduction of the *GBSSI* RNAi constructs. Waxy potato starch is expected to have clarity and stability, especially stability against the freeze-thaw treatment. These properties have been desired in the food

Table 28.2 Metabolic engineering of crop plants by RNAi

		RNAi cassette			
Plant	Target gene	Target region	Promoter	Modified trait	References
Rice	OsBP-5	Coding region	Ubiquitin	Reduced amylose content	Zhu et al. (2003)
Rice	qSW5	Coding region	CaMV 35S	Increased seed weight	Shomura et al. (2008)
Wheat	1Dx5	5′ UTR*	Ubiquitin	Altered quality of flour	Yue et al. (2008)
Wheat	SBEIIa	Exons 1 ~ 3	HMW glutenin	Increased amylose content	Regina et al. (2006)
Maize	22-kD α-zein	5′ terminal cDNA region	27-kD γ-zein, zp22/6	High lysine kernel	Segal et al. (2003)
Maize	19- and 22-kD α-zein	Coding region	27-kD γ-zein	High lysine kernel	Huang et al. (2006b)
Maize	Lysine-ketogulutarate reductase	Coding region	Endosperm-specific promoter	High lysine kernel	Houmard et al. (2007)
Potato	SBE1 and SBE2	3′ terminal cDNA region	GBSSI	Increased amylose content	Andersson et al. (2006)
Potato	GBSSI	5′ coding region, Middle coding region 3′ coding and UTR region	GBSSI	Reduced amylose content	Heilersig et al. (2006)
Sweet potato	GBSSI	Exon 1	CaMV 35S	Reduced amylose content	Otani et al. (2007)
Soybean	GmFAD3	Coding region	Glycinin	Reduced α-linolenic acid content	Flores et al. (2008)
Tomato	TDET1	Coding region	Fruit-specific promoters	Increased carotenoid content	Davuluri et al. (2005)
Tomato	LTPG1, LTPG2	Coding region	CaMV 35S	Reduced allergen	Le et al. (2006a)
Tomato	Lyc e 1	Coding region	CaMV 35S	Reduced allergen	Le et al. (2006b)
Peanut	Ara h 2	Coding region	CaMV 35S	Reduced allergen	Dodo et al. (2008)
Cotton	ghSAD-1	Coding region	Lectin	Increased stearic acid content	Liu et al. (2002)
Cotton	ghFAD2-1	5′ terminal cDNA region	Lectin	Increased oleic acid content	Liu et al. (2002)
Cotton	δ-cadinene synthase	Coding region	α-globulin	Reduced toxic gossypol content	Sunilkumar et al. (2006)
Coffee	CaMXMT1	3′ coding and UTR region	CaMV 35S	Reduced caffeine content	Ogita et al. (2003)
Apple	Mal d 1	5′ UTR and exon 1	CaMV 35S	Reduced allergen	Gilissen et al. (2005)

UTR, untranslated region.

industry and in paper manufacture (Jobling 2004). Heilersig and colleagues made eight different RNAi constructs that contained different regions, lengths and orientation of potato *GBSSI* cDNA fragment, and transcription of inverted repeats was controlled by the *GBSSI*'s own promoter. They used a part of *GBSSI* cDNA as a spacer. The results showed that sense–antisense or antisense–sense orientations in the inverted repeat had nearly the same effects on silencing efficiency. High silencing efficiencies were observed when the constructs using 5′ region or middle region, but not 3′ region, of *GBSSI* cDNA were introduced into potato plants. The constructs with a stem size of 500 ~ 600 bp and a spacer of about 150 bp in length more efficiently silenced than the larger stems with a larger spacer. Although strong reduction in amylose core can be obtained, a complete reduction of *GBSSI* mRNA was never observed in the resulting transgenic potato plants. Otani et al. (2007) showed the effective elimination of the sweet potato (*Ipomoea batatas* L.) GBSSI protein by RNAi. Over 70% of the transgenic plants with an RNAi construct targeting the 351-bp-first exon of the *GBSSI* gene produced amlyose-free storage roots. This amylose-free starch showed altered physicochemical properties.

28.4.1.2 High-Amylose Starch

High-amylose cornstarch has been commercially produced, where naturally occurring mutants of maize (*Zea mays*) are cultivated. High-amylose starch is unique in its film forming, gelling properties, and resistance against digestion in the human gastrointestinal tract. This 'resistant starch' has nutritional benefits as dietary fibers (Topping et al. 2003). Until now, no other major crops with the high amylose starch have been found. Andersson et al. (2006) have succeeded in producing of high-amylose potato plants by simultaneous inactivation of two genes encoding starch branching enzymes (*SBE1* and *SBE2*). The 200-bp fragments of both *SBE* genes were fused, and a single inverted repeat cassette was prepared. This RNAi cassette was inserted behind a potato *GBSS* promoter. The resulting transgenic potato plants showed a high-amylose phenotype with amylose contents ranging from 38% to 87% whereas the amylose content of parental plants was about 20%. In monocots, three major isoforms of the starch branching enzyme have been identified, namely *SBEI*, *SBEIIa* and *SBEIIb*. In maize, a high-amylose phenotype is caused by a mutation in the *SBEIIb* gene (Kim et al. 1998). Regina et al. (2006) generated high-amylose wheat (*Triticum aestivum*) by RNAi-based technology. They showed that *SBEIIa* hairpin RNAs reduced the expression of both *SBEIIa* and *SBEIIb* genes, although the target region of *SBEIIa* is only 70% identical to the corresponding region of *SBEIIb*. There is one region in which a 21-nt long identical continuous stretch exists between *SBEIIa* and *SBEIIb* genes. In contrast, a hairpin construct against the *SBEIIb* gene did not inactivate the *SBEIIa* gene. The similarity between the *SBEIIb* target region and the corresponding *SBEIIa* gene sequences is about 70% and one 21-nt long stretch is found. The precise reason was unclear why the RNAi construct against the *SBEIIa* gene can reduce the expression of *SBEIIb* while

the opposite cannot. Transgenic wheat plants produced starch with >70% amylose content. The analysis of total starch content in the endosperm revealed a slight reduction in the transgenic wheat plants (43%) compared with 52% in the control plants. The high-amylose wheat grain improved the digestion of rats, indicating that it served as a source of resistant starch.

28.4.2 Improvement of Seed Oils

Vegetable oils produced from oilseed crops contain oleic (18:1), linoleic (18:2), α-linolenic (18:3) acids and it has been recommended that these oils be preferentially used in the human diet for beneficial nutritional effects. In contrast, polyunsaturated fatty acids (18:2 and 18:3) are too unstable for use in cooking. Stable cooking oils with reduced polyunsaturated fatty acids and an increased saturated fatty acid, stearic acid (18:0), are produced by an industrial process of hydrogenation. However, partial hydrogenation often produces *trans*-fatty acids that raise human health risks by increasing of plasma low-density lipoprotein (LDL). Thus, improvement of seed oils with reduced polyunsaturated fatty acids is an important subject of plant genetic engineering (Napier 2007).

Major enzymes involved in the fatty acid desaturation (FAD) have been cloned. The stearoyl-acyl-carrier protein Δ9 desaturase (SAD) converts the 18:0 into the 18:1 fatty acid. The 18:1 fatty acid is unsaturated by a Δ12 desaturase (FAD2) to produce the 18:2. The ω3 (equal to Δ15) fatty acid desaturase (FAD3) is responsible for the 18:3 synthesis from the 18:2. RNAi technology is apparently valuable for control of fatty acid composition in membrane lipids (Tomita et al. 2004; Hirai et al. 2007), and also in storage seed oils. Flores et al. (2008) showed that RNAi against the soybean (*Glycine max*) *FAD3* gene can reduce the 18:3 content by up to 85%. Soybean has three *FAD3* paralogs, namely *GmFAD3A*, *GmFAD3B*, and *GmFAD3C*. A 318-bp conserved domain of the *FAD3* coding region was used as an RNAi target region under the control of soybean glycinin seed-specific promoter. In transgenic soybean plants, these three *FAD3* genes were effectively silenced by a single RNAi construct, even though these paralogs have a similarity at sequence level of 85% to 96% identity in the targeted region.

Liu et al. (2002) prepared two RNAi constructs for silencing *SAD* (*ghSAD-1*) and *FAD2* (*ghFAD2-1*) genes of cotton plants (*Gossypium hirsutum*). Each construct contained a spacer sequence consisting of its own exon sequences and a seed-specific soybean lectin promoter. In developing seeds of the resultant transgenic cotton, expression levels of *ghSAD-1* and *ghFAD2-1* genes were markedly reduced. By silencing the *ghSAD-1* gene, the 18:0 content increased from the non-transformant level of 2% up to 40%, which was associated with a corresponding decrease of the 18:1 and 18:2 contents. In the transgenic cotton plants introduced with a *ghFAD2-1* RNAi construct, the 18:1 content increased up to 77% of total fatty acids at maximum while the corresponding content in control seeds was 15%.

Seed oils, especially palmitic acid (16:0), are also important as a starting material for various industrial applications. RNAi-based engineering of *Arabidopsis* plants with extremely high 16:0 contents is reported (Pidkowich et al. 2007).

28.4.3 Manipulation of Storage Proteins

Zeins, major maize seed storage proteins, are deficient in lysine content. Lysine is a limiting essential amino acid for animal nutrition. From analyses of naturally occurring maize mutants with high lysine contents, reduction of the zein content is considered to be a potent approach for increasing lysine content. A decrease of zein proteins should be associated with an increase of non-zein, lysine-rich proteins. Segal et al. (2003) made two RNAi constructs to reduce the 22-kD α-zein. One construct had an inverted repeat harboring the target zein cDNA fragment under the control of the endosperm-specific 27-kD γ-zein promoter. In this construct, sense and antisense fragments were directly linked. Another one had a green fluorescent protein (*GFP*) spacer sequence and its inverted repeat cassette was transcribed by a strong endosperm-specific promoter of the zp22/6 zein gene. Both resulting transgenic maize plants showed reduced accumulation of 22-kD α-zein, and the lysine content of kernels markedly increased. Huang et al. (2006b) produced transgenic maize plants with reduced accumulation of both 19- and 22-kD α-zeins. Two distinct RNAi constructs (pMON73566 and pMON73567) were prepared, in which the inverted repeat cassette was transcribed by the 27-kD γ-zein promoter. pMON73566 has an inverted repeat against the 19-kD zein gene, and antisense 22-kD zein sequences are included as a loop of the resulting hairpin RNAs. Introduction of pMON73566 resulted in low silencing frequency of both 19- and 22-kD zein genes. pMON73567 is comprised of an inverted repeat that contains tandem fused sequences with the 19- and 22-kD zein genes, and also harbors a spacer consisting of an unpaired 22-kD zein gene fragment. pMON73567 transformants showed an efficient reduction in the expression of both 19- and 22-kD zein genes. Total amino acid analysis demonstrated that two times higher levels of lysine and tryptophan were found in the pMON73567 kernels than those of parental plants. As an alternative approach to breeding of high-lysine corn, Houmard et al. (2007) reported the RNAi-mediated suppression of a maize lysine degradation enzyme, lysine-ketoglutarate reductase/saccharopine dehydrogenase. The hairpin RNAs were specifically expressed in endosperm, and lysine content in transgenic maize kernels significantly increased.

Wheat flour has been used in a wide range of distinct foods, and gluten proteins containing high molecular weight glutenin subunit (HMW-GS) are the most important component affecting cooking quality (Anjum et al. 2007). The contents and relative proportion of HMW-GS are closely involved in determination of dough performance, namely its strength and elasticity. Hexaploid wheat has five HMW-GS genes called 1Ax2, 1Dx5, 1Bx7, 1By9, and 1Dy10 on the homoeologous chromosomes 1A, 1B, and 1D. Yue et al. (2008) reported the RNAi inhibition of 1Dx5

expression. The RNAi construct comprised of inverted repeat sequences harboring a 200-bp fragment of the 1Dx5 gene, spacer sequences from the intron 4 of wheat *waxy* gene, and the maize ubiquitin promoter. Target sequences of the 1Dx5 gene showed 81–91% identity to the corresponding sequences of 1Ax2, 1Bx7, 1By9, and 1Dy10 genes. There was no >20-nt long perfect identity on target regions. Wheat transformation was conducted by particle bombardment. Analysis of HMG-GS expression in the resultant transgenic plants showed that the 1Dx5 expression was completely silenced. Furthermore, expression of the 1Bx7 gene was reduced in spite of no difference in the protein level of 1Ax2, 1By9, and 1Dy10 compared to the control plants. Analysis of flour quality showed that wheat flour from transgenic plants could not produce dough necessary to make bread because of a reduction of gluten development. Therefore, the genetic control of gluten components is possible by RNAi (for undesired HMG-GS gene) and overexpression (for desirable HMW-GS gene), which may develop new quality and taste of wheat flour.

28.4.4 RNAi-Mediated Reduction of Plant Allergens

Decrease of food allergen contents is one of the most important resolutions for avoidance of human hypersensitive reaction such as anaphylaxis. The typical plant food allergens can be classified into the well-known four protein families; prolamin, cupin, profilin, and Bet v 1 (Radauer and Breiteneder 2007). Prolamins comprise of dominant seed storage proteins of cereals, such as maize, wheat and rice. Albumins, α-amylase/trypsine inhibitors, and non-specific lipid transfer proteins (LTPs) are included in the prolamin superfamily. The cupin superfamily can be identified by sequence similarities between wheat germin, fungal spherulins, and plant globulins. Profilins are relatively small cytosolic proteins found in all eukaryotic cells. The Bet v 1 superfamily includes a family termed *pathogenesis-related proteins 10* (*PR-10*). Several transgenic plants with reduced allergens have been reported.

The IgE reactivities found in tomato (*Lycopersicon esculentum*) fruits are due to profilins (Lyc e 1), β-fructofuranosidase (Lyc e 2), a non-specific LTP (Lyc e 3) and other minor allergens. Le et al. (2006a, b) produced several transgenic tomato plants with reduced allergens. They identified two paralog genes encoding Lyc e 3, *LTPG1* and *LTPG2*. When an RNAi cassette specific to the *LTPG1* gene was expressed by CaMV 35S promoter, both *LTPG1* and *LTPG2* genes were efficiently silenced in the transgenic tomato fruits. The allergenic potential test using the *Lyc e 3*-diminished fruits revealed that histamine release from human basophiles strongly decreased compared with that seen in control tomato fruits (Le et al. 2006a). In addition, they reduced the tomato minor allergen Lyc e 1 content by RNAi technology (Le et al. 2006b). In the latter case, an RNAi vector was prepared by using pK7GWIWG2(II) vector (Karimi et al. 2002). Profilin accumulation in resulting transgenic tomato fruits decreased to about 10% of that seen in control tomato fruits.

Peanut (*Arachis hypogaea*) allergy is one of the most life-threatening food allergies. Many peanut allergens have been identified, namely, 7S globulin (Ara h 1), 2S albumin (Ara h 2), 11S storage globulin (Ara h 3/4), profilin (Ara h 5), minor 2S albumin (Ara h 6 and Ara h 7), Bet v 1-related protein (Ara h 8), lectin and oleosin. Dodo et al. (2008) produced transgenic peanut plants in which the *Ara h 2* gene was silenced. An RNAi vector harboring 265-bp long *Ara h 2* coding sequences was constructed by using pHANNIBAL vector (Wesley et al. 2001). ELISA and western immunoblot analysis in protein extracts from pod-grown transgenic peanuts showed that Ara h 2 content was reduced to 21~25% of that of control peanuts. The analysis using sera from five patients allergic to peanut revealed that the IgE binding capacity of transgenic peanut seeds markedly decreased compare to that of control seeds. It remains to be clarified whether or not these transgenic peanuts can effectively avoid anaphylaxis under clinical testing on the patients. Since most allergy patients are polysensitized to several plant allergens, it is hoped for food safety that lots of researchers try to produce hypoallergenic plants by RNAi targeting multiple allergens.

28.4.5 Engineering of Secondary Metabolism by RNAi

Plants produce so many kinds of secondary metabolites. Some of them are essential for human nutrition, and other metabolites such as alkaloids have a wide variety of uses because of their pharmacological effects on humans. RNAi technology has been applied for modulation of secondary metabolism in order to diminish or increase target metabolites.

Lycopene and β-carotene are members of carotenoids, and an adequate intake of these substances can prevent human diseases including cardiovascular diseases, cancer and other chronic diseases (Rao and Rao 2007). Carotenoids have antioxidant properties and can reduce damage by reactive oxygen species. β-carotene is a precursor of vitamin A. Deficiency of vitamin A causes blindness and growth retardation of children. Davuluri et al. (2005) showed that lycopene and β-carotene contents increased in transgenic tomato fruits by silencing of the *DET1* gene. *DET1*, *DE-ETIOLATED 1*, is well known as a negative regulator of light-controlled gene expression in *Arabidopsis* plants. The mutation in a tomato *DET1* homolog gene (*TDET1*) is responsible for the *hp-2* phenotype. The *hp-2* mutants display elevated levels of flavonoids and carotenoids in deeply pigmented fruits (Levin et al. 2003). An inverted repeat harboring the *TDET1* sequences was inserted behind three fruit-specific promoters, and resulting three RNAi constructs were introduced into tomato plants. These constructs effectively reduced the *TDET1* mRNA in tomato fruits, which was associated with significant increases of carotenoid and flavonoid contents. β-carotene contents of the transgenic tomato fruits showed more than a 5-fold increase compared to those of control fruits.

Caffeine (1,3,7-trimethylxanthine) is a purine alkaloid, and is synthesized during seed development and in leaves of coffee and tea plants. Caffeine has physiological

effects on humans such as central nervous system stimulation, increased blood pressure and diuretic effect (Higdon and Frei 2006). A recent increase of consumption of coffee and caffeine has become a problem as it raises a health risk. Decaffeinated coffee is produced by an industrial process, but it loses important coffee flavor. For production of coffee plants with reduced caffeine contents, RNAi technology has been applied. Theobromine, a precursor of caffeine, is synthesized by methylation of 7-methylxanthine. This enzymatic step is catalyzed by 7-N-methylxanthine methyltransferase (MXMT) (Ashihara et al. 2008). Ogita et al. (2003) succeeded in silencing the expression of a coffee MXMT gene (*CaMXMT1*). The coffee (*Coffea canephora*) plants were transformed with an RNAi construct targeting the 3′ untranslated region of *CaMXMT1* gene. The caffeine content of young leaves of resulting transgenic plants was reduced to 50–70% of that of control leaves. However, the caffeine content in coffee beans has not yet been reported. They also produced caffeine-reduced Arabica coffee plants (*Coffea arabica*) by using the same RNAi construct (Ogita et al. 2004).

28.5 Viral Resistance by RNAi

Since RNA silencing functions as a part of the defense mechanism against viruses, dsRNA-mediated plant viral resistance was demonstrated at an early phase of RNA silencing studies (Waterhouse et al. 1998; Smith et al. 2000). Until now, successful immunity against plant viruses was obtained in several transgenic plants expressing hairpin RNAs harboring viral sequences (e.g., Kalantidis et al. 2002; Di Nicola-Negri et al. 2005; Kamachi et al. 2007) and amiRNAs targeting viral sequences (Niu et al. 2006; Qu et al. 2007). To counteract the RNA silencing defense mechanism, plant viruses have RNA silencing suppressors in their genomes (Silhavy and Burgyán 2004). Many kinds of suppressors have been identified, but no obvious sequence homology has been found among distinct suppressors. Some viral suppressors bind siRNAs, and sequester siRNAs from host RNA silencing machineries (Vargason et al. 2003; Ye et al. 2003; Goto et al. 2007; Lózsa et al. 2008). Could these viral suppressors inhibit the action of RNAi against the viral sequences? In most cases, viral infection can be inhibited in plants expressing dsRNAs harboring the cognate viral sequences. Cucumber mosaic virus (CMV) contains the silencing suppressor gene, 2b. The transgenic plants expressing dsRNAs against the CMV coat protein gene were resistant to CMV infection, indicating that CMV 2b cannot suppress the host RNAi against CMV (Kalantidis et al. 2002). In addition, transgenic plants expressing the dsRNAs specific to the viral silencing suppressor gene showed the almost complete immunity (Di Nicola-Negri et al. 2005; Niu et al. 2006; Qu et al. 2007).

As mentioned above, RNAi can efficiently generate resistance against RNA viruses. However, it is rather difficult to make stable RNAi transgenic plants immune to DNA viruses (Poogin and Hohn 2004). Only a few reports showed possibility of RNAi-mediated resistance to DNA viruses. For example, siRNAs corresponding to a replication-associated protein gene of a geminivirus, African

cassava mosaic virus (ACMV), were co-inoculated with ACMV, and a significant reduction in viral DNA accumulation was observed (Vanitharani et al. 2003). Since plant DNA viruses also express silencing suppressor proteins (e.g., begomovirus AC2, Voinnet et al. 1999), RNAi against suppressor genes (Ramesh et al. 2007) and/or inactivation of viral transcription by RdDM are now considered (Poogin and Hohn 2004).

As simultaneous infection with several kinds of viruses is common in the field, broad virus resistance is a desirable trait. However, due to the high sequence specificity, RNAi has been limited to the inhibition of a single virus infection. Bucher et al. (2006) developed an RNAi construct harboring multiple trigger sequences in a single inverted repeat structure. The resultant transgenic plants were resistant to four different tospoviruses. Niu et al. (2006) produced a transgenic plant that expressed two distinct amiRNAs from a dimeric amiRNA transgene. These two amiRNAs targeted two distinct plant viruses, turnip yellow mosaic virus and turnip mosaic virus. As expected, the resultant plants were resistant to both viruses. The broadness of viral resistance is also expected to be extended in future.

28.6 Control of Plant-Feeding Pests by Host RNAi

The finding of RNAi in *C. elegans* was followed by extensive investigation of its mechanism. One such early work showed that feeding *E. coli* expressing the target dsRNAs is sufficient to induce RNAi (Timmons and Fire 1998). Therefore, dsRNAs in the gut should systemically spread into the somatic tissues and germ line, which open a possibility that dsRNAs are delivered from plants to the plant-feeding pests.

28.6.1 Nematode-Resistant Crops

Plant-parasitic nematodes cause extensive damage to crops, and a $125 billion worth of crop losses is estimated (Fuller et al. 2008). The use of chemical nematicides is effective in reducing the level of soil nematodes, but their cost and toxicity to humans impose financial loads and health risks on growers. Plant-parasitic nematodes are generally obligate plant parasites, and they have a stylet, needle-like organ, that is used to pierce the plant cells for injection of nematode secretions and feeding plant cytoplasm (Gheysen and Vanholme 2006). Recently RNAi-based approaches were reported to control nematode infestation (Huang et al. 2006a, Yadav et al. 2006). A parasitism gene, *16D10,* of root-knot nematodes encodes a secretory peptide that stimulates root growth and functions as a ligand for a putative plant transcription factor. RNAi-mediated reduction of *16D10* mRNA level in the nematode resulted in reduced nematode infectivity. In fact, *Arabidopsis* plants expressing a dsRNA specific to *16D10* showed enhanced resistance against 4 major

root-knot nematodes (Huang et al. 2006a). Yadav et al. (2006) introduced RNAi vectors harboring target sequences for a nematode splicing factor gene and an integrase gene into tobacco plants. The resultant plants were protected against nematode infection. The mRNA levels for the splicing factor and integrase genes markedly decreased in the nematodes, but mRNA levels of other nematode genes did not significantly change (Yadav et al. 2006). These results indicate that inactivation of target genes in nematodes probably results from an RNAi-like mechanism.

28.6.2 Insect-Resistant Crops

The systemic silencing in *C. elegans* requires amplification of siRNAs. In *C. elegans*, siRNAs annealed to a target mRNA serve as primers for RNA-Dependent RNA Polymerase (RDR), and a complementary RNA molecule is synthesized. The newly synthesized dsRNAs are subjected to the Dicer-dependent processing, and secondary siRNAs (transitive siRNAs) are generated (Sijen et al. 2001). This amplification process contributes to the long-lived effects of RNAi in *C. elegans* (Tabara et al. 1998). Unlike the case of *C. elegans*, RNA silencing in *Drosophila melanogaster* is restricted in the site of dsRNA delivery and is temporally limited. A key component of systemic RNA silencing is RDR, but a gene encoding an RDR-related protein has not been identified in *D. melanogaster* (Roignant et al. 2003). Therefore, some researchers wonder if dsRNAs in the diet can diffuse into somatic tissues and silence the target genes in insect larvae. Two transgenic plants producing dsRNAs against insect essential genes were recently produced and these plants are resistant to insect pests. Thus, the finding obtained by using *D. melanogaster* would not be always applicable to the RNAi mechanisms of other insects.

Cotton bollworm (*Helicoverpa armigera*) can grow with a diet containing high concentrations of a cotton metabolite, gossypol. *CYP6AE14*, a gossypol-inducible gene, is involved in the *H. armigera* tolerance against gossypol. When larvae were reared on transgenic tobacco and *Arabidopsis* leaves expressing dsRNAs specific to *CYP6AE14*, the *CYP6AE14* mRNA level in larvae was reduced to undetectable levels 7 days after feeding. Then these larvae were transferred to artificial diets containing gossypol. Larvae with decreased level of *CYP6AE14* mRNA did not grow at all (Mao et al. 2007).

Western corn rootworm (*Diabrotica virgifera virgifera* LeConte), a coleopteran species, causes root damage. To select target genes, Baum et al. (2007) carried out a feeding assay, in which larvae were fed with artificial diets containing various dsRNA species. A gene encoding V-type ATPase was selected. When the transgenic maize plants expressing a dsRNA specific to the V-type ATPase gene were infested with Western corn rootworm eggs, these plants showed a significant reduction in worm feeding damage.

These two examples indicate that RNAi can be exploited to control insect pests by *in planta* expression of a dsRNA specific to the target insect gene. In contrast to *D. melanogaster*, a systemic spread of the RNAi signal from the gut should occur

in the cotton bollworm and western corn rootworm. The red flour beetle, *Tribolium castaneum*, responds to dsRNA systemically, and its complete genomic sequence is now available (*Tribolium* genome sequencing consortium 2008). As was the case of *D. melanogaster*, a homolog gene for an *RDR* is deficient in the genome of *T. castaneum*, suggesting that systemic RNA silencing in *T. castaneum* is likely to be significantly different from that in *C. elegans*, and still remains to be uncovered (Tomoyasu et al. 2008).

28.7 Caveats and Future Perspectives of RNAi Technology

It is possible that siRNAs down-regulate mRNA species other than the intended target mRNA. This side effect is called 'off-target effect' (Jackson et al. 2003). Xu et al. (2006) investigated the off-target effects in the plant RNAi, and they reported that sequence identity with at least 22-nt continuous stretches is required for silencing of unintended genes. When the transcriptional changes in RNAi-mediated silencing and non-silencing backgrounds were monitored by using microarray techniques, no significant transcriptional changes were detected between these two samples (Aelbrecht et al. 2006). These results imply that changes in the transcriptome are the direct result of the knockdown of target genes in most cases. The validation of 'on-target' silencing is apparently easy when amiRNA technology is used for knockdown of target genes. In the latter case, we can prepare two or more different target sequences per single target gene. 'On-target' silencing should cause the same silencing phenotypes among transformants with different amiRNAs.

RNAi and related silencing machineries serve as plant defense against invasive nucleic acids, such as transposons (Lippman et al. 2003; Kanno et al. 2005) and viruses (Voinnet 2005). In the course of plant transformation, siRNAs corresponding to T-DNA sequences are produced (Canto et al. 2002). This observation suggests that delivery of T-DNA into plant cells itself is recognized as a target of RNAi-related defense machineries. In fact, the *RDR6*-deficient plants and plants expressing a viral suppressor, P38, are significantly more susceptible to *Agrobacterium*-mediated transformation than control plants (Dunoyer et al. 2006). *RDR6* is one of the six paralogs of *RDR* genes in *Arabidopsis* plants, and plays a pivotal role in the sense-transgene induced RNA silencing and also in the formation of secondary siRNAs (Schwach et al. 2005). It is worthwhile to test whether or not the RNAi against the host *RDR6* homolog gene improves *Agrobacterium*-mediated transformation efficiency in the transformation-recalcitrant crops. The *RDR6*-deficiency also enhances accumulation of recombinant proteins (Butaye et al. 2004). One notable drawback is that lack of *RDR6* is associated with increased susceptibility with plant viruses (Qu et al. 2005; Schwach et al. 2005). Therefore, the transgenic plants that are planned to be commercially released must have the functional *RDR6* gene. Finally, the unique properties of siRNAs, namely mobile ability between cells and also between organisms in some cases, opens room for development of new applications that have not been expected.

References

Aelbrecht T, Vuylsteke M, Bauwens M et al (2006) Introduction of silencing-inducing transgenes does not affect expression of known transcripts. FEBS Lett 580:4154–4159

Andersson M, Melander M, Pojmark P et al (2006) Targeted gene suppression by RNA interference: an efficient method for production of high-amylose potato lines. J Biotechnol 123:137–148

Anjum FM, Khan MR, Din A et al (2007) Wheat gluten: high molecular weight glutenin subunits-structure, genetics, and relation to dough elasticity. J Food Sci 72:R56-R63

Ashihara H, Sano H, Crozier A (2008) Caffeine and related purine alkaloids: biosynthesis, catabolism, function and genetic engineering. Phytochemistry 69:841–856

Baulcombe D (2004) RNA silencing in plants. Nature 431:356–363

Baum JA, Bogaert T, Clinton W et al (2007) Control of coleopteran insect pests through RNA interference. Nat Biotechnol 25:1322–1326

Baumberger N, Baulcombe DC (2005) *Arabidopsis* ARGONAUTE1 is an RNA slicer that selectively recruits microRNAs and short interfering RNAs. Proc Natl Acad Sci U S A 102:11928–11933

Bernstein E, Caudy AA et al (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 409:363–366

Brodersen P, Sakvarelidze-Achard L, Bruum-Rasmussen M et al (2008) Widespread translational inhibition by plant miRNAs and siRNAs. Science 320:1185–1190

Bucher E, Lohuis D, van Poppel PMJA et al (2006) Multiple virus resistance at a high frequency using a single transgene construct. J Gen Virol 87:3679–3701

Butaye KMJ, Goderis IJWM, Wouters PFJ et al (2004) Stable high-level transgene expression in *Arabidopsis thaliana* using gene silencing mutants and matrix attachment regions. Plant J 39:440–449

Canto T, Cillo F, Palukaitis P (2002) Generation of siRNAs by T-DNA sequences does not require active transcription or homology to sequences in the plant. Mol Plant Microbe Interact 15:1137–1146

Cheung F, Haas BJ, Goldberg SMD et al (2006) Sequencing *Medicago truncatula* expressed sequenced tags using 454 Life Sciences technology. BMC Genomics 7:272

Chuang CF, Meyerowitz EM (2000) Specific and heritable genetic interference by double-stranded RNA in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A 97:4985–4990

Davuluri GR, van Tuinen A, Fraser PD et al (2005) Fruit-specific RNAi-mediated suppression of *DET1* enhances carotenoid and flavonoid content in tomatoes. Nat Biotechnol 23:890–895

Di Nicola-Negri E, Brunetti A et al (2005) Hairpin RNA-mediated silencing of *Plum pox virus* P1 and HC-pro genes for efficient and predictable resistance to the virus. Transgenic Res 14:989–994

Dodo HW, Konan KN, Chen FC et al (2008) Alleviating peanut allergy using genetic engineering: the silencing of the immunodominant allergen Ara h 2 leads to its significant reduction and a decrease in peanut allergenicity. Plant Biotechnol J 6:135–145

Douchkov D, Nowara D et al (2005) A high-throughput gene-silencing system for the functional assessment of defense-related genes in barley epidermal cells. Mol Plant Microbe Interact 18:755–761

Dunoyer P, Himber C, Voinnet O (2006) Induction, suppression and requirement of RNA silencing pathways in virulent *Agrobacterium tumefaciens* infections. Nat Genet 38:258–263

Dunoyer P, Himber C, Ruiz-Ferrer V et al (2007) Intra- and intercellular RNA interference in *Arabidopsis thaliana* requires components of the microRNA and heterochromatic silencing pathways. Nat Genet 39:848–856

Fire A, Xu S, Montgomery MK et al (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391:806–811

Flores T, Karpova O, Su X et al (2008) Silencing of Gm*FAD3* gene by siRNA leads to low α-linolenic acids (18:3) of *fad3*-mutant phenotype in soybean [*Glycine max* (Merr.)]. Transgenic Res 17:839–850

Fuller VL, Lilley CJ, Urwin PE (2008) Nematode resistance. New Phytol 180:27–44

Fusaro AF, Matthew L, Smith NA et al (2006) RNA interference-inducing hairpin RNAs in plants act through the viral defence pathway. EMBO Rep 7:1168–1175
Gheysen G, Vanholme B (2006) RNAi from plants to nematodes. Trends Biotechnol 25:89–92
Gilissen LJ, Bolhaar ST, Matos CI et al (2005) Silencing the major apple allergen Mal d 1 by using the RNA interference approach. J Allergy Clin Immunol 115:364–369
Goto K, Kobori T, Kosaka Y et al (2007) Characterization of silencing suppressor 2b of Cucumber mosaic virus based on examination of its small RNA-binding abilities. Plant Cell Physiol 48:1050–1060
Hamilton AJ, Baulcombe DC (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 286:950–953
Heilersig HJB, Loonen A, Bergervoet M et al (2006) Post-transcriptional gene silencing of GBSSI in potato: effects of size and sequence of the inverted repeats. Plant Mol Biol 60:647–662
Helliwell C, Waterhouse P (2003) Constructs and methods for high-throughput gene silencing in plants. Methods 30:289–295
Higdon JV, Frei B (2006) Coffee and health: a review of recent human research. Crit Rev Food Sci Nutr 46:101–123
Hilson P, Allemeersch J, Altmann T et al (2004) Versatile gene-specificsequence tags for *Arabidopsis* functional genomics: transcript profiling and reverse genetics applications. Genome Res 14:2176–2189
Himmelbach A, Zierold U, Hensel G et al (2007) A set of modular binary vectors for transformation of cereals. Plant Physiol 145:1192–1200
Hirai S, Kodama H (2008) RNAi vectors for manipulation of gene expression in higher plants. Open Plant Sci J 2:31–40
Hirai S, Oka S et al (2007) The effects of spacer sequences on silencing efficiency of plant RNAi vectors. Plant Cell Rep 26:651–659
Houmard NM, Mainville JL, Bonin CP et al (2007) High-lysine corn generated by endosperm-specific suppression of lysine catabolism using RNAi. Plant Biotechnol J 5:605–614
Huang G, Allen R, Davis E et al (2006a) Engineering broad root-knot resistance in transgenic plants by RNAi silencing of a conserved and essential root-knot nematode parasitism gene. Proc Natl Acad Sci U S A 103:14302–14306
Huang S, Frizzi A, Florida CA et al (2006b) High lysine and high tryptophan transgenic maize resulting from the reduction of both 19- and 22-kD α-zeins. Plant Mol Biol 61:525–535
Jackson AL, Bartz SR, Schelter J et al (2003) Expression profiling reveals off-target gene regulation by RNAi. Nat Biotechnol 21:635–637
Jobling S (2004) Improved starch for food and industrial applications. Curr Opin Plant Biol 7:210–218
Kalantidis K, Psaradakis S et al (2002) The occurrence of CMV-specific short RNAs in transgenic tobacco expressing virus-derived double-stranded RNA is indicative of resistance to the virus. Mol Plant Microbe Interact 15:826–833
Kamachi S, Mochizuki A et al (2007) Transgenic *Nicotiana benthamiana* plants resistant to cucumber green mottle mosaic virus based on RNA silencing. Plant Cell Rep 26:1283–1288
Kanno T, Huettel B, Mette MF et al (2005) Atypical RNA polymerase subunits required for RNA-directed DNA methylation. Nat Genet 37:761–765
Karimi M, Inzé D, Depicker A (2002) GATEWAY vectors for *Agrobacterium*-mediated plant transformation. Trends Plant Sci 7:193–195
Kim KN, Fisher DK et al (1998) Molecular cloning and characterization of the *Amylose-Extender* gene encoding starch branching enzyme IIB in maize. Plant Mol Biol 38:945–956
Le LQ, Lorenz Y, Scheurer S et al (2006a) Design of tomato fruits with reduced allergenicity by dsRNAi-mediated inhibition of ns-LTP (Lyc e 3) expression. Plant Biotechnol J 4:231–242
Le LQ, Mahler V, Lorenz Y et al (2006b) Reduced allergenicity of tomato fruits harvested from Lyc e 1-silenced transgenic tomato plants. J Allergy Clin Immunol 118:1176–1183
Levin I, Frankel P, Gilboa N et al (2003) The tomato *dark green* mutation is a novel allele of the tomato homolog of the *DEETIOLATED1* gene. Theor Appl Genet 106:454–460

Lippman Z, May B, Yordan C et al (2003) Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. PLoS Biol 1:420–428
Liu Q, Singh SP, Green AG (2002) High-stearic and high-oleic cottonseed oils produced by hairpin RNA-mediated post-transcriptional gene silencing. Plant Physiol 129:1732–1743
Lózsa R, Csorba T et al (2008) Inhibition of 3' modification of small RNAs in virus-infected plants require spatial and temporal co-expression of small RNAs and viral silencing-suppressor proteins. Nucleic Acids Res 36:4099–4107
Mao Y, Cai W, Wang J et al (2007) Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. Nat Biotechnol 25:1307–1313
Matzke M, Kanno T, Huettel B et al (2007) Targets of RNA-directed DNA methylation. Curr Opin Plant Biol 10:512–519
Mi S, Cai T, Hu Y et al (2008) Sorting of small RNAs into *Arabidopsis* Argonaute complexes is directed by the 5' terminal nucleotide. Cell 133:116–127
Miki D, Shimamoto K (2004) Simple RNAi vectors for stable and transient suppression of gene function in rice. Plant Cell Physiol 45:490–495
Mlotshwa S, Pruss GJ, Peragine A et al (2008) *DICER-LIKE2* plays a primary role in transitive silencing of transgenes in *Arabidopsis*. PLoS ONE 3:e1755
Napier JA (2007) The production of unusual fatty acids in transgenic plants. Annu Rev Plant Biol 58:295–319
Niu QW, Lin SS, Reyes JL et al (2006) Expression of artificial microRNAs in transgenic *Arabidopsis thaliana* confers virus resistance. Nat Biotechnol 24:1420–1428
Ogita S, Uefuji H, Yamaguchi Y et al (2003) Producing decaffeinated coffee plants. Nature 423:823
Ogita S, Uefuji H et al (2004) Application of RNAi to confirm theobromine as the major intermediate for caffeine biosynthesis in coffee plants with potential for construction of decaffeinated varieties. Plant Mol Biol 54:931–941
Ossowski S, Schwab R, Weigel D (2008) Gene silencing in plants using artificial microRNAs and other small RNAs. Plant J 53:674–690
Otani M, Hamada T, Katayama K et al (2007) Inhibition of the gene expression for granule-bound starch synthase I by RNA interference in sweet potato plants. Plant Cell Rep 26:1801–1807
Pidkowich MS, Ngyen HT, Heilmann I et al (2007) Modulating seed β-ketoacyl-acyl carrier protein synthase II level converts the composition of a temperate seed oil to that of a palm-like tropical oil. Proc Natl Acad Sci U S A 104:4742–4747
Pooggin MM, Hohn T (2004) Fighting geminiviruses by RNAi and vice versa. Plant Mol Biol 55:149–152
Pop M, Salzberg SL (2007) Bioinformatics challenges of new sequencing technology. Trends Genet 24:142–149
Qu F, Ye X, Hou G et al (2005) RDR6 has a broad-spectrum but temperature-dependent antiviral defense role in *Nicotiana benthamiana*. J Virol 79:15209–15217
Qu J, Ye J, Fang R (2007) Artificial microRNA-mediated virus resistance in plants. J Virol 81:6690–6699
Radauer C, Breiteneder H (2007) Evolutionary biology of plant food allergens. J Allergy Clin Immunol 120:518–525
Ramesh SV, Mishra AK, Praveen S (2007) Hairpin RNA-mediated strategies for silencing of *tomato leaf curl virus* AC1 and AC4 genes for effective resistance in plants. Oligonucleotides 17:251–257
Rao AV, Rao LG (2007) Carotenoids and human health. Pharmacol Res 55:207–216
Regina A, Bird A, Topping D et al (2006) High-amylose wheat generated by RNA interference improves indices of large-bowel health in rats. Proc Natl Acad Sci U S A 103:3546–3551
Roignant J, Carré C, Mugat B et al (2003) Absence of transitive and systemic pathways allows cell-specific and isoform-specific RNAi in *Drosophila*. RNA 9:299–308
Schwach F, Vaistij FE et al (2005) An RNA-dependent RNA polymerase prevents meristem invasion by potato virus X and is required for the activity but not the production of a systemic silencing signal. Plant Physiol 138:1842–1852

Segal G, Song R, Messing J (2003) A new opaque variant of maize by a single dominant RNA-interference-inducing transgene. Genetics 165:387–397
Shomura A, Izawa T, Ebana K et al (2008) Deletion in a gene associated with grain size increased yields during rice domestication. Nat Genet 40:1023–1028
Sijen T, Fleenor J, Simmer F et al (2001) On the role of RNA amplification in dsRNA-triggered gene silencing. Cell 107:465–476
Silhavy D, Burgyán J (2004) Effects and side-effects of viral RNA silencing suppressors on short RNAs. Trends Plant Sci 9:76–83
Smith NA, Singh SP, Wang MB et al (2000) Total silencing by intron-spliced hairpin RNAs. Nature 407:319–320
Song J, Smith SK et al (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. Science 305:1434–1437
Sönnichsen B, Koski LB, Walsh A et al (2005) Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. Nature 434:462–469
Sunilkumar G, Campbell LM, Puckhaber L et al (2006) Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. Proc Natl Acad Sci U S A 103: 18054–18059
Tabara H, Grishok A, Mello CC (1998) RNAi in *C. elegans*: soaking in the genome sequence. Science 282:430–431
Timmons L, Fire A (1998) Specific interference by ingested dsRNA. Nature 395:854
Tomita R, Hamada T, Horiguchi G et al (2004) Transgene overexpression with cognate small interfering RNA in tobacco. FEBS Lett 573:117–120
Tomoyasu Y, Miller SC, Tomita S et al (2008) Exploring systemic RNA interference in insects: a genome-wide survey for RNAi genes in *Tribolium*. Genome Biol 9:R10
Tribolium genome sequencing consortium (2008) The genome of the developmental model beetle and pest *Tribolium castaneum*. Nature 452:949–955
Topping DL, Fukushima M, Bird AR (2003) Resistant starch as a prebiotic and synbiotic: state of the art. Proc Nut Soc 62:171–176
Vanitharani R, Chellappan P, Fauquet CM (2003) Short interfering RNA-mediated interference of gene expression and viral DNA accumulation in cultured cells. Proc Natl Acad Sci U S A 100:9632–9636
Vargason JF, Szittya G et al (2003) Size selective recognition of siRNA by an RNA silencing suppressor. Cell 115:799–811
Voinnet O (2005) Non-cell autonomous RNA silencing. FEBS Lett 579:5858–5871
Voinnet O, Pinto YM, Baulcombe DC (1999) Suppression of gene silencing: a general strategy used by diverse DNA and RNA viruses of plants. Proc Natl Acad Sci U S A 96:14147–14152
Waterhouse PM, Graham MW, Wang MB (1998) Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. Proc Natl Acad Sci U S A 95:13959–13964
Wesley SV, Helliwell CA, Smith NA et al (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. Plant J 27:581–590
Whitehurst AW, Bodemann BO, Cardenas J et al (2007) Synthetic lethal screen identification of chemosensitizer loci in cancer cells. Nature 446:815–819
Wielopolska A, Townley H, Moore I et al (2005) A high-throughput inducible RNAi vector for plants. Plant Biotechnol J 3:583–590
Xie Z, Johansen LK, Gustafson AM et al (2004) Genetic and functional diversification of small RNA pathways in plants. PLoS Biol 2:642–652
Xu P, Zhang Y, Kang L et al (2006) Computational estimation and experimental verification of off-target silencing during posttranscriptional gene silencing in plants. Plant Physiol 142:429–440
Yadav BC, Veluthambi K, Subramaniam K (2006) Host-generated double stranded RNA induces RNAi in plant-parasitic nematodes and protects the host from infection. Mol Biochem Parasitol 148:219–222

Ye K, Malinina L, Patel DJ (2003) Recognition of small interfering RNA by a viral suppressor of RNA silencing. Nature 426:874–848

Yu B, Yang Z, Li J et al (2005) Methylation as a crucial step in plant microRNA biogenesis. Science 307:932–935

Yue SJ, Li H, Li YW et al (2008) Generation of transgenic wheat lines with altered expression levels of 1Dx5 high-molecular weight glutenin subunit by RNA interference. J Cereal Sci 47:153–161

Zheng X, Zhu J et al (2007) Role of *Arabidopsis* AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. EMBO J 26:1691–1701

Zhu Y, Cai XL et al (2003) An interaction between a MYC protein and an EREBP protein is involved in transcriptional regulation of the rice *Wx* gene. J Biol Chem 278:47803–47811

Zilberman D, Cao X, Johansen LK et al (2004) Role of *Arabidopsis ARGONAUTE4* in RNA-directed DNA methylation triggered by inverted repeats. Curr Biol 14:1214–1220

Chapter 29
Metabolomics in Fruit Development

Kati Hanhineva and Asaph Aharoni

Abstract Metabolomics aims at the efficient determination of multiple chemical constituents present in a tissue, a cell layer or ideally a single cell. Metabolomics is currently applied in a large number of life science disciplines. Nuclear Magnetic Resonance (NMR) or Gas- and Liquid-Chromatography coupled to Mass Spectrometry (GC-MS and LC-MS) are the most widespread technologies employed in metabolomics assays. Soft fruit, one of the most metabolite-rich plant organs, was from the first to be subjected to metabolomics investigation. The interest in metabolite profiling of soft fruit could be explained by the large repertoire of metabolites belonging to diverse chemical classes that are formed and catabolized during fruit development, starting from the fertilized ovary up to the ripe, mature fruit. Moreover, fruit constitute an essential part of our diet and the breeding to achieve nutrient-rich varieties entails a comprehensive analysis of their metabolite content. Presently, hundreds of substances, including primary and secondary (or specialized) metabolites have been detected in fruit. Metabolomics has been employed also for following metabolism in transgenic plants, mutants and introgression lines populations. The latter experiments allowed the identification of genomic regions associated with metabolic quality traits. So far, most metabolomics assays in fruit have been focused on two species, namely, tomato and strawberry. It is expected that in the following years the use of metabolomics will be expanded to the investigation of numerous other fruit species.

K. Hanhineva
The University of Kuopio, Kuopio, Finland

A. Aharoni (✉)
The Weizmann Institute of Science, Rehovot, Israel
e-mail: asaph.aharoni@weizmann.ac.il

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_29,

29.1 Introduction

Metabolomics is one of the most rapidly advancing analytical approaches that aim at the comprehensive analysis of a large numbers of metabolites. It is currently applied in numerous fields of research, from drug discovery to brain function and food quality. It is most extensively utilized in the plant science field, particularly since plants possess the unique feature of synthesizing a large number of natural products that do not exist in any other life form. The current estimation of the number of different compounds produced in the plant kingdom exceeds 200,000 (Saito et al. 2006; Hall 2005). Natural products (also known as secondary metabolites) participate in the interaction of the plant with the environment, and serve multiple tasks in plant reproduction as for example in alluring pollinators and frugivores. Several important attributions for human health also originate from plant metabolites including the impact on food quality and nutritional value (e.g. carotenoid content; Fraser and Bramley 2006) or drug research (e.g. taxol cell suspension culture; Ketchum and Croteau 2006). Therefore, metabolomics analysis of plant-based foodstuff like fruit, berries and cereals, currently gains a great interest.

Fruit constitute a variety of different tissues and make a most complex part of the plant physiology. As part of our food, fruit serve as an important source of fibre, vitamins and other phytochemicals, therefore being the target of several studies that concentrate on food quality and nutritional value. In addition to mature fruit, also the preceding and later stages of fruit development, the flowers and the seeds, produce a myriad of bioactive phytochemicals that may serve as health-beneficial or phytopharmacological source. Therefore, the study of the reproductive developmental process is of major interest not only fundamentally for learning about the mechanisms and interplay between different chemicals, but also from an applied point of view.

In the first part of this chapter we describe the basic principles and methodologies of the current metabolomics approaches and their application in the plant science field. In the second part we review the current research on soft fruit development utilizing metabolomics. Here, we focus on strawberry and tomato fruit since they have been the two major targets for metabolite profiling efforts since the onset of plant metabolomics research.

29.2 Metabolomics in Plant Science

29.2.1 What Is Metabolomics?

The detection of metabolites has been carried out for decades, the technical landmark being the development of chromatographic separation methods that enabled the detection of individual metabolites (Unger 2004; Fritz 2004). Traditionally, metabolite analysis has been performed in a targeted manner in which a single or several compounds, normally structurally and biologically associated, were examined.

Several issues promoted the development and use of metabolite analysis approaches that carry out a non-targeted analysis of metabolites in a given biological matrix (Seger and Sturm 2007; Villas-Boas et al. 2005). These include the increased amount of predicted new metabolic pathways due to genome sequencing and the emergence of systems biology analysis of metabolic networks (Ryan and Robards 2006; Fiehn et al. 2000). Today, the term metabolomics is used in parallel with transcriptomics and proteomics that are focused on mRNA and protein profiling, respectively. Metabolomics is therefore defined as the systematic quantitative and qualitative study of the complement of metabolites in a particular biological system. Among the different "omics" approaches, metabolomics is the most predictive since the metabolome represents the final measurable response of an organism (Seger and Sturm 2007; Ryan and Robards 2006). A typical metabolomics workflow includes several steps starting from sample extraction, through data pre-prosessing and statistical analysis, metabolite identification and finally the determination of the biological impact (Fig. 29.1).

29.2.2 Analytical Platforms used In Metabolomics Assays

The significant breakthroughs in the development of analytical tools for metabolite analysis paved the way for the simultaneous monitoring of a multitude of small molecules that are essential for metabolomics experiments. The pros and cons and details regarding the capabilities of the different technologies for metabolomics were described in many reviews (e.g. Seger and Sturm 2007; Dettmer et al. 2007; Glinski and Weckwerth 2006; Dunn et al. 2005; Villas-Boas et al. 2005). Mass-spectrometry (MS)-based technologies that permit the measurement of the masses of individual ions are currently the most commonly used detection methods for small molecules. MS allows the sensitive detection and identification of metabolites through the interpretation of mass spectrum and determination of molecular formulae *via* accurate mass measurements (Dunn et al. 2005).

The analytical instruments typically used in metabolomics experiments differ in their ionization technology (e.g. electron ionization (EI), chemical ionization (CI), electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), matrix-assisted laser desorption/ionization (MALDI) and fast atom/ion bombardment (FAB)), and the type of mass analyzer ((e.g. quadrupole, triple-quadrupole, ion-trap and time of flight (TOF)) (Villas-Boas et al. 2005). Samples can be directly analyzed by MS or resolved primarily by different chromatographic techniques. The choice of analytical method depends on which cellular metabolites are the targets of analysis.

Gas-Chromatography-MS (GC-MS) has a long history as a metabolite detection technology and it provides advantage in terms of robustness (Fernie et al. 2004). This technology has high separation efficiency that can resolve complex volatile mixtures that include alcohols, aldehydes, furans, ketones, terpenes among other low molecular weight compounds, which may be collected by e.g. solvent free headspace sampling. GC-MS is further used in metabolomics technologies for the detection of non-volatile metabolites through the chemical derivatization of the

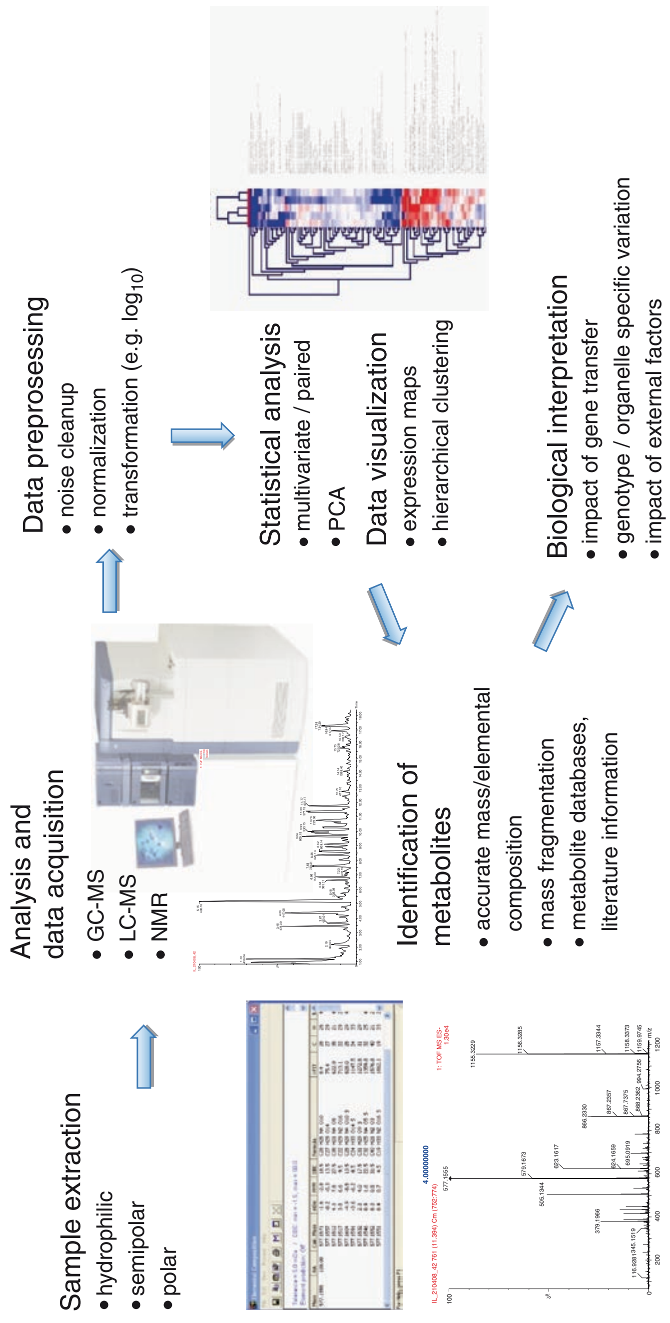


Fig. 29.1 A scheme of the metabolomics workflow

extract analytes in order to render them amenable for analysis. Common derivatization procedures include trimethylsilylation and the reaction with diazomethane (Halket et al. 2005, Seger and Sturm 2007, Dettmer et al. 2007).

Coupling Liquid-Chromatography to MS (LC-MS) enables the detection of multiple metabolite classes in a single analysis, even in a very complex matrix such as rich plant extracts (Dunn et al. 2005). The technology for liquid phase separation have also advanced in recent years with the introduction of Ultra Performance LC (UPLC) that enables faster separation with better separation and sensitivity as compared to High Performance LC (HPLC). Capillary-electrophoresis-mass-spectrometry (CE-MS) is a highly sensitive method that can detect low-abundance metabolites with good analyte separation (Villas-Boas et al. 2005; Dunn et al. 2005). Accurate mass measurements are an important factor in metabolomics experiments since they allow improved structural elucidation. Fourier-transform-ion-cyclotron-resonance-mass-spectrometry (FTICR-MS), a different type of mass analyzer, permits mass determination at the highest accuracy and supreme resolution of mass spectra beyond any other instrument (Brown et al. 2005). Nuclear-magnetic-resonance spectroscopy (NMR) is one of the most widely applied methods to study organic molecules, and has been used for over 20 years for metabolite analysis. NMR is a non-destructive method that is highly versatile and adjustable for any chemical analysis. It is the ultimate tool to unravel the identity of a compound, as each separate resonance observed in a spectrum is specific to a particular atom pair, and the careful translation of individual signals leads to the identification of any component regarded as a target (Ward et al. 2007; Krishnan et al. 2005; Dunn et al. 2005).

Despite the extensive development in analytical methods the ultimate goal of plant metabolomics, i.e. to gain a complete overview of the metabolite complement of a plant in one or a small series of analyses, is currently impossible. The continuous technological improvements, including the use of complex hyphenated systems as for example the combination of NMR and LC-MS may bring the field to its ultimate goal as described above (Hall 2005; Ryan and Robards 2006).

29.2.3 Metabolomics Data Processing and Its Mining in a Biological Context

The multiple experimental platforms applied in metabolomics research generate complex data matrices that require dedicated software and laborious handling. These have a big impact on the outcome of metabolomics experiments and on their biological relevance (Katajamaa and Orešič 2005, 2007). For example, a typical LC-MS raw dataset typically contains information on retention time, *m/z* value and ion intensity measurement for each signal. Several processing steps including filtering, feature alignment and detection, and normalization are mostly applied in order to filter and condense the data set. The data is subsequently processed using diverse mining tools including multivariate analyses and clustering for discovering differences between groups of samples (Katajamaa and Orešič 2007). Thus, bioinformatics and biostatistics

are among the most essential disciplines utilized in metabolome research. Automated computational tools are currently developed that allow metabolite identification, quantification and the detection of differential metabolites. The markedly increased amounts of published metabolomics data and the desire to maintain the data accessible and comparable between laboratories urged the development of a minimal metabolomics data reporting standards (Fiehn et al. 2007). The same developments promoted the construction of metabolomics databases as for example KNApSAcK (Shinbo et al. 2006) and MoTo (Moco et al. 2006).

An even bigger challenge is faced by researchers who attempt to link metabolomics data to other biological observations including those achieved by equivalent large-scale "omics" tools such as expression arrays and proteomics. The integration of information of both bottom-up approaches representing the traditional way of trying to understand the complex biological pathways by reducing the information to the straight interactions of its parts (the "reductionistic" view) as well as top-down approaches; to take into account the possibility that the whole system most likely possess more features than only the straight interactions of its parts (the "holistic" view), are necessary to construct hierarchical networks to understand any biological system (Lange 2006; Morgenthal et al. 2006; Damian et al. 2007).

29.3 Metabolomics Applications in Fruit Development

29.3.1 The Process of Fruit Development

In the majority of species fruit development commences with pollination and extends via ripening of the fruit to the maturation of the seeds. The first stage of fruit development, the formation of the inflorescence, has been the target of intensive research for the past 20 years. At anthesis, floral organs form the entity responsible for plant reproduction via pollen development, attraction of pollinators, fertilization and induction of fruit and seed development. The next step after fertilization is characterized by organ expansion and formation of the seeds. Once the fruit has reached its final size and weight, ripening, the terminal stage of fruit development begins and a major shift in metabolism occurs. In addition to physiological changes such as the modification of cell wall structure and texture, the ripening of fruit is accompanied by the accumulation of aroma components as well as sugars (to provide attracting flavor and taste), and protective agents such as phenolic compounds that serve as anti-pathogens and UV-protectants (Gillaspy et al. 1993; Bewley and Black 1994). The fundamentals of fruit development and its genetic basis have been mainly studied in the dry, dehiscent fruit of the model plant Arabidopsis (Robles and Pelaz 2005), while species producing soft fruit have been the main targets of interest regarding maturation and the metabolic changes associated with it.

In the area of metabolite characterization of fruit, the studies on two cultivated fruit crops, strawberry and tomato, dominate the field. These two species offer two highly distinct fruit structures both physiologically and biochemically, and hence,

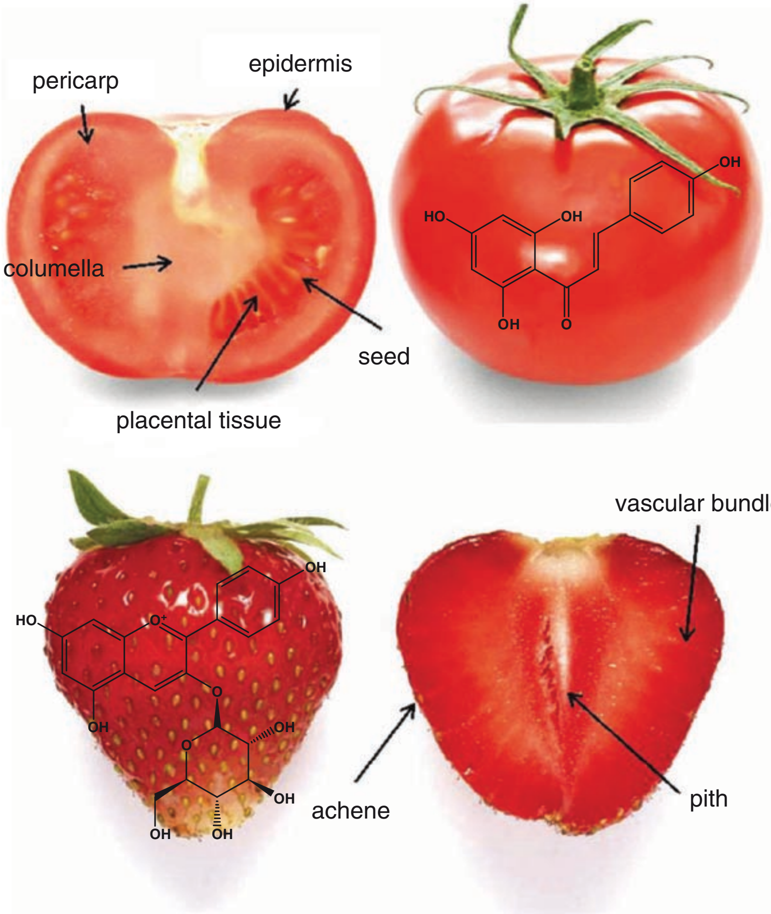


Fig. 29.2 Cross-section of tomato and strawberry fruit. The most characteristic tissue types are marked for each fruit. The chemical structure of typical secondary metabolites found in each fruit are depicted, naringenin chalcone for tomato and pelargonidin glucoside in the case of strawberry

represent unique fruit development and maturation programs. While tomato is a true fruit having a peel tissue enclosing the fleshy pericarp and seeds, strawberry is termed "false fruit", with the achenes (i.e. the "seeds") developing on the surface of the swollen receptacle tissue (i.e. the "fruit"; see Fig. 29.2.). The maturation process of these two fruits also show major differences, as tomato is a climacteric fruit in which maturation is an ethylene-dependent process. During ripening climacteric fruit like tomato, banana, apple and avocado, show increased respiration and ethylene biosynthesis and emission. In contrast, non-climacteric fruit including strawberry and several citrus species do not posses the typical climacteric respiration and autocatalytic ethylene formation. Ethylene is required for the coordination and completion of ripening in climacteric fruit, whereas it is not yet associated in the developmental program of

non-climacteric fruit, although they may respond to ethylene as well (Giovannoni 2007). In the secondary metabolism of the two fruit species, the major shift is related to the formation of an intensive color as well as flavor and aromatic volatiles. In tomato, the photosynthetic membranes of chloroplasts are degraded, as they transit to chromoplasts and serve for the synthesis of carotenoids including β-carotene and lycopene. In strawberry, accumulation of anthocyanins result in the intensive red color. In addition to formation of pigment-related compounds, also other branches of secondary metabolism are rearranged, as the ratio of different metabolite classes vary along development and maturation for both species.

29.3.2 Metabolomics in Tomato Fruit Development

29.3.2.1 Technical Advances Exemplified in Tomato Fruit

Tomato is one of the most widely cultivated fruit crop that also serves as a model for research on fleshy fruit development (Giovannoni et al. 1995, Giovannoni 2007), including for metabolic studies (Carrari and Fernie 2006). To date, metabolomics in tomato fruit was carried out by both GC-MS and high-resolution LC-MS technologies. Following LC-MS analysis that was coupled to NMR for absolute identification of selected metabolites, Moco et al. (2006) described the establishment of a database dedicated to LC-MS-based information on tomato fruit metabolites. In the MoTo database, the information from ESI-qTOF-MS was used for the assignment of mass spectral peaks, based on the accurate mass, retention time, UV-spectral characteristics, and MS/MS fragmentation data. The open access database contained information on typical tomato metabolites including polyphenolic compounds and alkaloids and is therefore most valuable for other studies utilizing similar analytical techniques. A recent report by Iijima Y, Nakamura Y, Ogata Y et al. (2008) Metabolite annotations based on the integration of mass spectral information. Plant J. 54:949–62 utilized FTICR-MS technology for profiling tomato fruit development. The authors developed a method for systematic analysis of the high accurate mass FTICR-MS measurements in order to elucidate and classify the entire array of metabolites present in the tomato fruit sample. All together, the chemical building blocks for 869 metabolites were reported, and a system for grading the level of metabolite identification as part of the metabolite annotation procedure was introduced that facilitates the interpretation of the metabolite data in a biological context.

In a different approach, GC-MS technology was used to profile fruit volatiles in a non-targeted manner (Tikunov et al. 2005). A multi-step strategy was employed that consisted of (i) spectral alignment of the GC-MS profiles, (ii) multivariate analysis of phenotypes at the level of molecular fragments, and (iii) mass spectral reconstruction allowing metabolite recognition and identification. A set of 94 tomato genotypes were profiled for fruit volatile compounds revealing the presence of 322 different volatiles. Multivariate data analysis including hierarchical tree clustering, Principal Component Analysis (PCA) and the construction of a metabolite-metabolite correlation matrix, were used for non-targeted data treatment and extracted the differences in the composition of volatile metabolites between the different genetic backgrounds.

29.3.2.2 Metabolite Changes During Tomato Fruit Development

The most dramatic change in tomato fruit development occurs during the transition to the ripening process that is coupled to the accumulation of numerous secondary metabolites. Moreover, the transition to ripening is accompanied by a massive change to metabolism as a result of the de-greening process, in which the photosynthetically active chloroplasts are differentiated to chromoplasts. In tomato, as in other plant species, the fruit is composed of several different tissues and cell layers. As exemplified recently for tomato fruit skin (or peel) and flesh, these tissues vary in terms of transcript and metabolic profiles (Moco et al. 2007; Mintz-Oron et al. 2008). Mintz-Oron et al. (2008) described a comparative transcriptome and metabolome analyses of peel and flesh tissues during five stages of tomato fruit development. Metabolite profiling by UPLC-qTOF-MS and GC-MS revealed 100 metabolites that were enriched in the peel tissue during development. Out of the 45 secondary metabolites identified that were at least twofold up-regulated in the peel compared with the flesh tissue, thirty metabolites were derived from the phenylpropanoid pathway. Apart from phenylpropanoids, a large number of glycoalkaloids that serve as antifungal agents were identified in tomato peel. These profiling experiments unravelled a yet undiscovered pathway of glycoalkaloid metabolism. It starts from α-tomatine that accumulates to very high concentrations in early fruit development and decreases in the mature stages, and ends with lycoperosides that display an opposite profile of accumulation during fruit development.

As a continuation of the work in which a tomato metabolome database was generated (Moco et al. 2006; see above), Moco et al. (2007) profiled secondary metabolites using several analytical platforms, in order to asses the tissue specificity of metabolites in tomato peel and flesh tissues (Moco et al. 2007). They analyzed different compound classes including carotenoids, xantophylls, chlorophylls, tocopherols, flavonoids, phenolic acids, glycoalcaloids and saponins together with several other secondary metabolites and few primary metabolites. The analysis was focused on the peel and flesh of several commercial tomato cultivars, and similar metabolite trends were observed regardless the different genetic background. An even more detailed analysis was performed on the different stages of development (i.e. green, breaker, turning, pink, and red fruit) of the tomato cultivar *Ever*. Different tissue types of each developmental stage, including the vascular attachment region, epidermis, pericarp, columella and placenta, and jelly parenchyma including the seeds, were separated and analysed.

Apart from changes to secondary metabolism, central carbon metabolism has also been studied in tomato fruit. Early metabolomics studies using GC-MS analyses of derivatized extracts reported on the characterization of major shifts in primary metabolites during tomato development (Roessner et al. 2001; Roessner-Tunali et al. 2003). The major changes were reported to occur in sugar metabolism, since the major hexoses, glucose and fructose, accumulated at the ripe stage, whereas sucrose, hexose phosphates and the majority of sugar alcohols decreased. Furthermore, upon ripening the levels of aromatic amino acids increased, while those of TCA cycle intermediates decreased.

29.3.2.3 Metabolomics of Tomato Fruit in Transgenic and Mutant Plants during Development

In recent years, transgenic and tomato fruit mutants were also subject for metabolite profiling. Extensive metabolite analysis was carried out on transgenic tomato plants expressing the yeast *S*-adenosylmethionine decarboxylase (*ySAMdc*) gene (Mattoo et al. 2006; Neelam et al. 2008). Apart from the accumulation of polyamines, the ripe fruit of transgenic plants exhibited higher fructose to glucose and sugar to acid ratios. The metabolic analyses involved assays of amino acids, sugars, organic acids, choline and a few other unidentified metabolites by NMR. These studies provided a good example for the utilization of NMR in metabolomics of transgenic fruit plants.

GC-MS and HPLC-MS techniques were applied in a study focusing on the metabolite levels of the light-hyperresponsive high-pigment (*hp*) tomato mutant plants (Bino et al. 2005). Several metabolites, including carotenoids and flavonoids, known for their antioxidant or photoprotective activities were overproduced in the fruit of transgenic lines and these results provided new information regarding the metabolic response of fruit to light stress. The increased levels of bioactive metabolites in the red ripe fruit detected by metabolomics suggested that these plants could be exploited as a potentially commercial product with enhanced nutritional value.

Fraser and colleagues carried out extensive metabolomics work on ripening tomato fruit mainly with respect to carotenoids (Fraser et al. 2000, 2007a, b). HPLC, MALDI-MS and GC-MS were employed for the assay of a transgenic line constitutively expressing an additional phytoene synthase-1 (*Psy-1*) gene, during five stages of fruit development and ripening (Fraser et al. 2007b). In the mature green stage, 54 components, most of them carotenoids, were increased, whereas in the ripe red stage only a few qualitative differences were detected between the wild type and the *Psy-1* overexpressing transgenic lines. Specific parts of metabolism were altered in the green stage by the perturbations in carotenoid biosynthesis resembling metabolism during fruit ripening (e.g. formation of chromoplast-like structures), although the general fruit developmental properties such as ethylene production and fruit firmness remained unaltered. This suggested that the changes to pigmentation, plastid type, and metabolism associated with *Psy*-1 overexpression are not connected with the ripening process (Fraser et al. 2007b).

29.3.2.4 Combining Tomato Genetics and Metabolomics

The genetic variation present in the current cultivated tomatoes represents only a fraction of the genetic resources present in wild species. One of the major targets of tomato genetics studies is the exploration of potential genetic markers that might be introduced into cultivated species for breeding desirable traits like nutritional and taste quality. Remarkable effort has been put on the analysis of the segregation of genetic regions responsible for the quantitative phenotypic variation, the quantitative trait loci (QTL; Lippman et al. 2007). Typically, the approach has been used

for identifying chromosomal regions responsible for economically important characteristics such as yield and firmness, but during the past years the focus was also devoted to the analyses of QTLs that influence the metabolite composition of fruit. Coupling the identification of genomic regions associated with valuable quality traits with the analysis of the metabolite content allows the determination of associations between phenotypes and metabolite profiles.

In a preliminary analysis to identify QTLs associated with quality-related traits (including organic acids, sugars and other biochemical factors), 1–27 putative QTLs were identified for each of the 15 traits analysed (Fulton et al. 2002). Soon after, a more comprehensive phenotypic analysis of the tomato introgression lines (ILs; containing a chromosome segment of wild species *Solanum pennellii* in the genetic background of the cultivated variety M82) was combined with parallel metabolite profiling by GC-MS (Schauer et al. 2006). The analysis of fruit pericarp tissue involved quantification and identification of 74 parimary metabolites affecting the tomato fruit quality including amino and organic acids, sugars, sugar alcohols and fatty acids. Correlation network and mapping analyses showed that at least 50% of the metabolic loci are associated with QTLs that modify yield associated traits. The study demonstrated that the whole-plant phenotypes correlate with fruit metabolism and hence have a significant role in the final metabolite composition of the fruit (Schauer et al. 2006). This work was continued with an analysis of samples from a second year harvest to determine the mode of inheritance for the QTLs related to tomato fruit metabolic traits (Schauer et al. 2008). The primary metabolites of the heterozygous ILs were profiled by GC-MS and 174 out of the 332 QTLs were shown to be dominantly inherited.

29.3.3 Metabolomics in Strawberry Fruit Development

The strawberry genus (*Fragaria*) belongs to the subfamily Rosoideae of the Rosaceae family that includes other known fruit species including apples, pears, plums, peaches and raspberries. Fruit of the Rosaceae family are known to have exceptionally rich secondary metabolite composition upon ripening. The ripe stage of strawberry development is well documented with respect to its metabolite composition (Koponen et al. 2007; Puupponen-Pimiä et al. 2005; Hannum 2004). The largest metabolite group consists of central phenylpropanoid and flavonoid pathway compounds including phenolic acid derivatives, flavonols, condensed tannins and anthocyanins. In addition to the commonly occurring phenolic compounds, remarkable amounts of ellagic acid and ellagitannins are present in plants of the Rosaceae family (Kähkönen et al. 2001), including in strawberry (Mullen et al. 2003; Hukkanen et al. 2007). Ellagitannins are polyphenols that are not products of the phenylpropanoid pathway, but rather generated through the shikimate pathway via the 5-dehydroshikimate precursor (Werner et al. 2004). The composition of strawberry ellagitannins varies between the receptacle tissue and the achenes (Aaby et al. 2005).

One of the early metabolomics experiments that employed FTICR technology was carried out in strawberry receptacle tissue (Aharoni et al. 2002). In this study,

the metabolite shifts in the strawberry receptacle were monitored along four consecutive stages of strawberry fruit development. Crude plant extracts were introduced via direct (continuous flow) injection and ionized by either electrospray ionization (ESI) or atmospheric pressure chemical ionization (APCI) in both positive or negative ionization modes. Changes in the levels of known fruit compounds, as well as putative compounds not previously associated with strawberry fruit development, revealed novel information on the metabolic transition from immature to ripe fruit. Integrating the data on metabolism during the early stages of strawberry fruit development revealed the accumulation of compounds that serve as raw material and building blocks for the production of ripening-associated metabolites. For example, the relative high levels of various organic, phenolic and amino acids in unripe fruit may serve as precursors for the formation of aromatic/phenolic components, while other amino acids such as glutamine may be used for the production of nitrogen-containing metabolites. Oxidative breakdown of the accumulated unsaturated fatty acids may result in the biosynthesis of aroma compounds (i.e. aldehydes, alcohols and esters). Carbon skeletons produced in the TCA cycle may be used to synthesise certain amino acids that, in turn, can be converted into more complex, secondary metabolites. A most important output of these experiments was that the method described could simultaneously and rapidly detect masses putatively corresponding to various types of compounds known to accumulate during ripening of strawberry fruit. Ripening associated metabolites that could be detected included flavour compounds such as furans, ester derivatives (aldehydes, alcohols and esters), terpenes, carbohydrates, anthocyanins and different other phenolics.

A recent, extensive survey of both primary and secondary metabolism in strawberry receptacle and achene tissues during development was carried out by combining LC-MS and GC-MS–based metabolite profiling (Fait et al. 2008). Six stages of development were studied, and the analysis highlighted a clear metabolic shift between the first three and the later three stages, as exemplified by the PCA of the masses obtained by the UPLC-qTOF-MS analysis (mainly corresponding to secondary metabolites) (Fig. 29.3). The level of each metabolites belonging to the same metabolite class and even different substituted forms of the same metabolite were tightly correlated both in the achenes and receptacle. The study corroborated earlier knowledge on primary metabolism, concluding that glucose, fructose and sucrose are the major soluble sugars in ripe strawberry (Hancock 2000), whereas inositol, xylose and galactose decrease in levels during ripening (Moing et al. 2001). The main organic acids detected were citrate, malate and quinate together with the minor organic acids acetate, oxalate, succinate, isocitrate, fumarate, and aconitate, which were also studied earlier in strawberry (Moing et al. 2001). The levels of sugar phosphates and sugar alcohols were also significantly reduced during receptacle development. Interestingly, the accumulation patterns of sugars in the achene tissue were in most cases different, and even opposite, from the ones observed in the receptacle.

The profiles of secondary metabolites in the two tissues showed that the phenolic compounds display the most noticeable metabolic changes during the development of strawberry fruit. The accumulation of the astringent condensed tannins characterizes

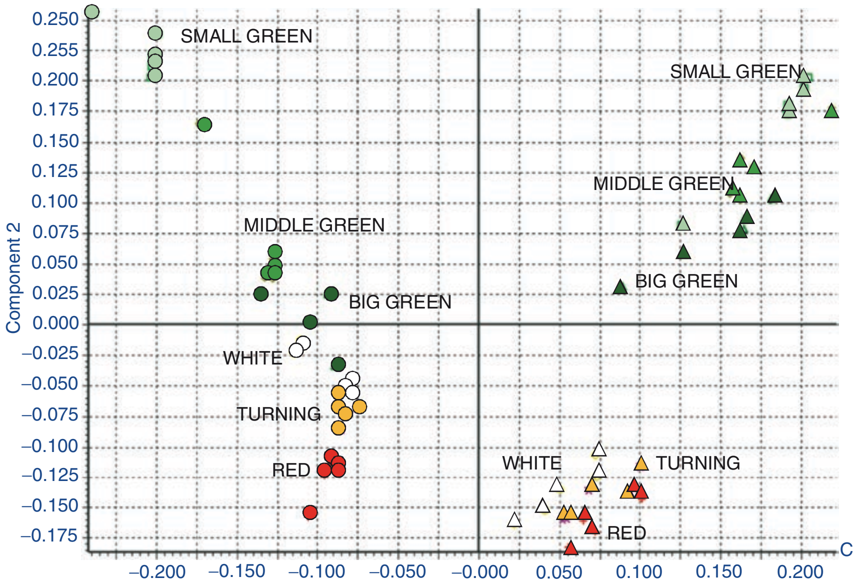


Fig. 29.3 Principal Component Analysis (PCA) of UPLC-QTOF-MS data set obtained by the profiling of strawberry fruit tissues during 6 stages of development. Circles and triangles are receptacle and achenes samples, respectively (n = 5 per stage)

the early stages of development. Upon ripening, the levels of these metabolites decrease, whilst flavonoids such as anthocyanidins and flavonols predominate the collection of phenolic compounds. The differences in phenolics constituents between the two tissue types were large. For example, different derivatized forms of flavonols showed distinct accumulation, as exemplified by the profile of kaempferol coumaroyl glucoside present mainly in the achene tissue (Fig. 29.4.). In addition, ellagitannins and their precursors (galloyl glucoses) showed a very peculiar accumulation in the two tissue types. While the single galloylated form was mainly detected in the receptacle, the highest galloylation degree (penta-galloyl-glucose) was present in the achenes (Fig. 29.5). The achene tissue also contained ample chromatographic peaks that possessed the typical ellagitannin spectra, but none have been characterized previously from strawberry.

29.3.4 Comparison of Tomato and Strawberry Metabolite Profiles During Development

The pattern of secondary metabolism in tomato and strawberry is highly distinct; this can already be observed in the total ion chromatogram obtained by the UPLC-qTOF-MS analysis of fruit from the two species (Fig. 29.6). The most pronounced difference between the two species is the absence of glycoalkaloids from straw-

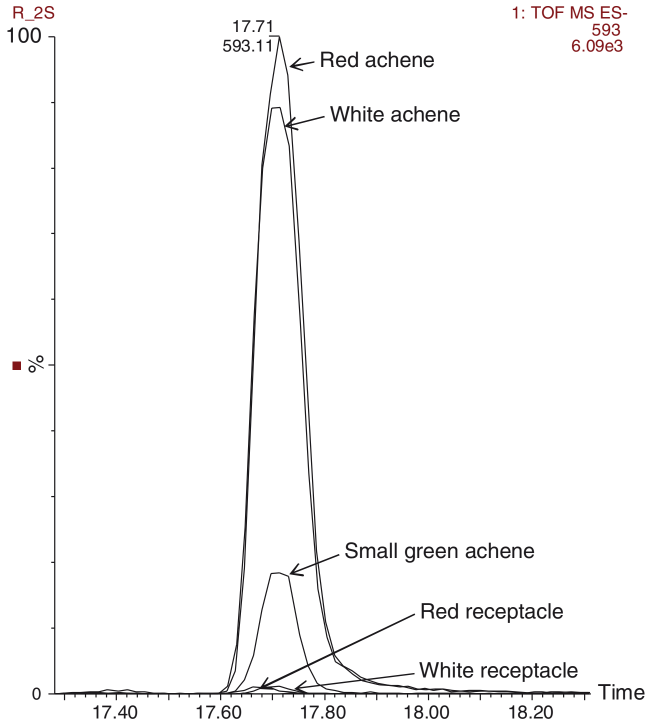


Fig. 29.4 Reconstructed ion chromatogram of the metabolite *m/z* 593 (ES−) identified as kaempferol coumaroyl glucoside. The chromatograms are shown on the same scale, and the developmental stage and tissue is indicated for each

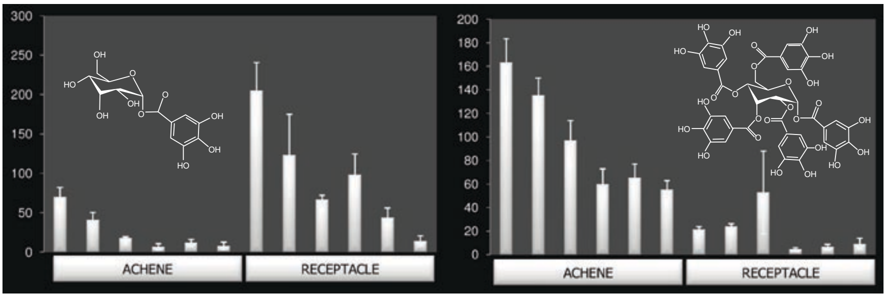


Fig. 29.5 Accumulation profiles of galloyl-glucose and penta-galloyl-glucose during strawberry fruit development. The average mass peak response area is shown on the same scale for both achene and receptacle, normalized to the dry weight. Six stages of development are shown, from left to right: small green, middle green, mature green, white, turning and red fruit. Galloyl-glucose (*right*) and penta-galloyl-glucose (*left*)

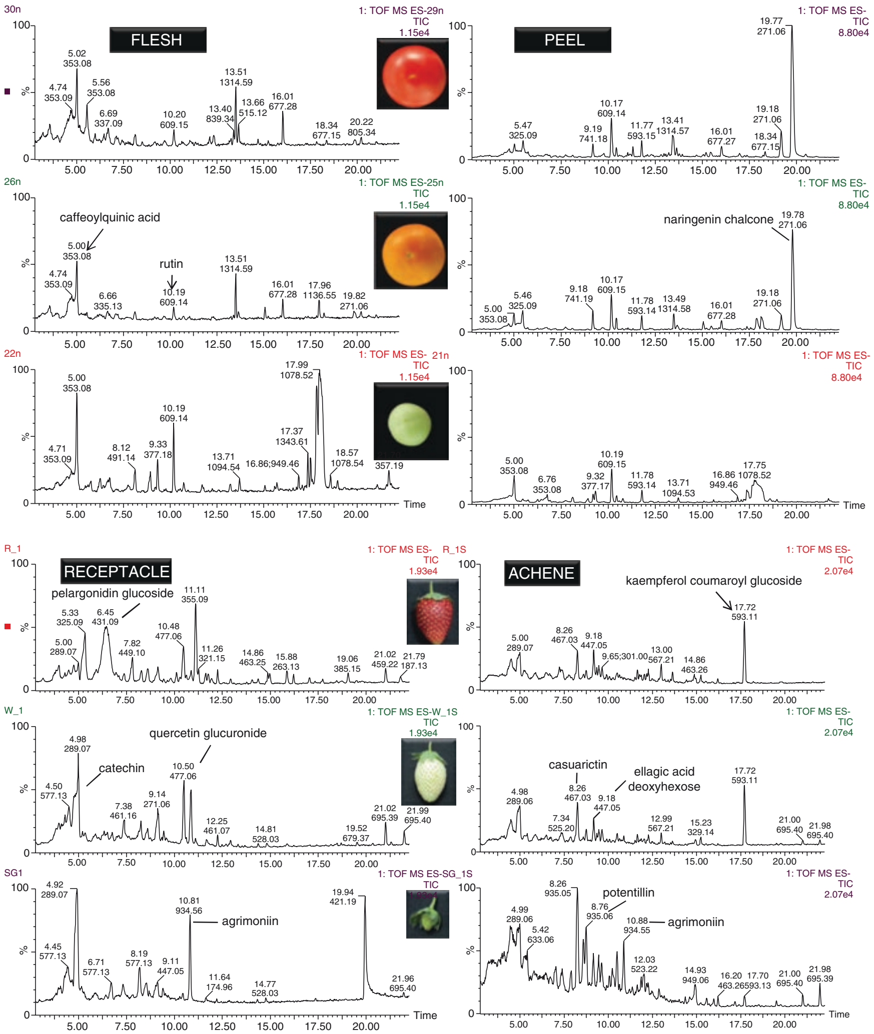
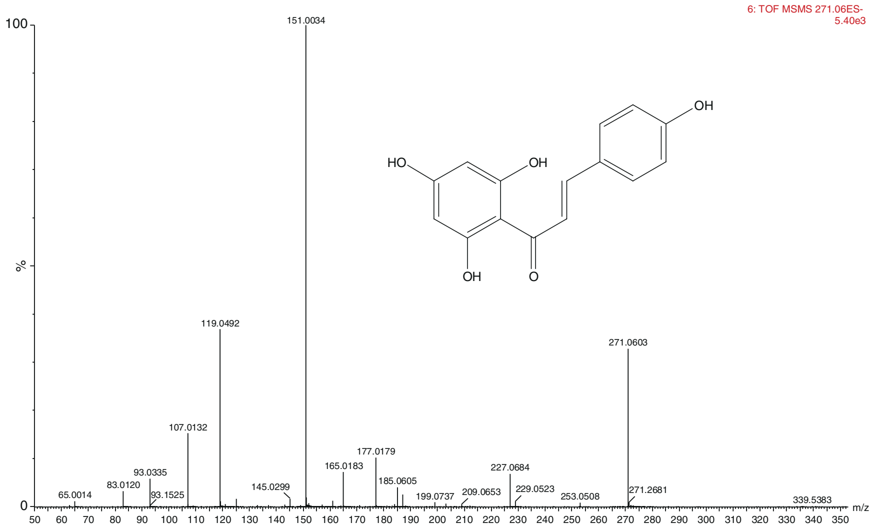


Fig. 29.6 Total ion chromatograms of strawberry and tomato fruit development obtained by UPLC-qTOF-MS analysis. The chromatograms of tomato (flesh and peel) and strawberry (receptacle and achenes) were obtained by analysing three stages of either fruit development

berry, while they dominate the metabolite profile of tomato. On the other hand, strawberry fruit contains a variety of ellagitannin derivatives, which are not synthesized in tomato. The outermost tissues of both tomato and strawberry show a very intensive accumulation of flavonoids. The achenes, developing on the strawberry receptacle surface, accumulate large amounts of a flavonol derivative, kaempferol coumaroyl glucoside (Fait et al. 2008; Figs. 29.4 and 29.7), while the yellow naringenin chalcone accumulates in tomato peel (Moco et al. 2007; Fig. 29.7).



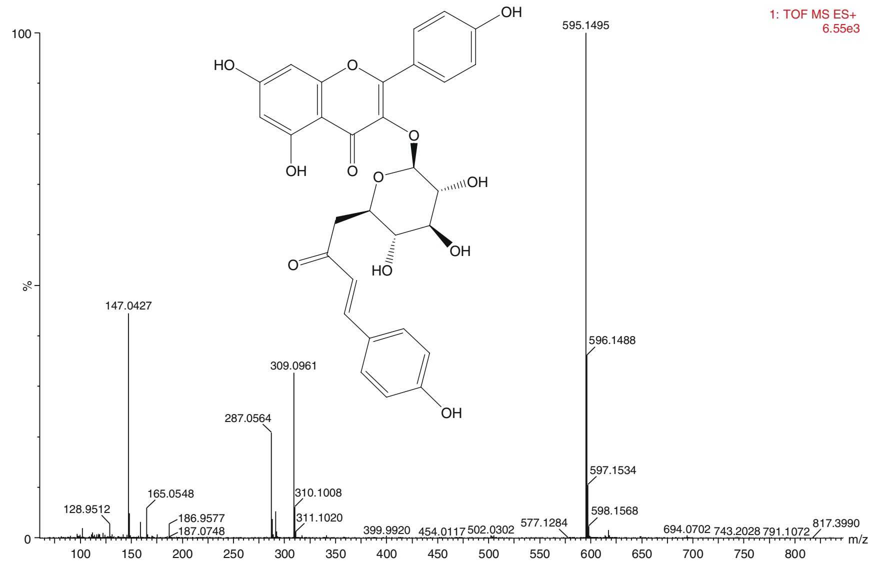


Fig. 29.7 The fragmentation patterns of flavonoids that accumulate to high levels in tomato and strawberry fruit. MS/MS spectra of naringenin chalcone (ES−), upper panel, and kaempferol coumaroyl glucoside (ES+), lower panel, are presented

The role for these metabolites is most likely to protect from UV-radiation, and possibly to serve as antifungal agents. The behavior of the primary metabolites is strikingly similar in the climacteric tomato fruit as well as in the non-climacteric strawberry. The soluble sugars, sugar phosphates and sugar alcohols are significantly reduced during strawberry receptacle and tomato flesh development.

29.4 Conclusions

Fruits are one of the most metabolite-rich organs of plants. During their development, starting from the expanding immature fruit and up to the late ripening stage, the complex network of genes, proteins and metabolites are drastically changed. Due to its ability to follow a relatively large number of metabolites in a single or a few analyses, metabolomics makes an excellent tool for following metabolism in developing fruit. Adding the commercial value of fruit products to the written above makes soft fruit an excellent model for metabolomics studies. While at present most studies employing metabolomics were focused on tomato and strawberry fruit, it is expected that in the near future similar studies will be carried out in various other fruit species. The use of metabolomics has demonstrated the huge diversity of secondary metabolites produced by fruit. This information will be essential for the more thorough investigation regarding the relation between the presence or absence of specific metabolites (or metabolite classes) and their function in fruit physiology. The results up to now also highlight the strong link between secondary and primary metabolism. Apart from extending metabolomics to other fruit species it is also expected that future studies will be able to follow metabolism in various fruit cell layers and even cell compartments.

References

Aaby K, Skrede G, Wrolstad RE (2005) Phenolic composition and antioxidant activities in flesh and achenes of strawberries (*Fragaria ananassa*). J Agric Food Chem 53:4032–4040

Aharoni A, DeVos CHR, Verhoeven HA et al. (2002) Nontargeted metabolome analysis by use of fourier transform ion cyclotron mass spectrometry. OMICS 6:217–234

Bewley JD, Black M (1994) Seeds: physiology of development and germination, Ed. 2. Plenum, NY

Bino RJ, deVos CHR, Lieberman M et al. (2005) The light-hyperresponsive *high pigment-2^{dg}* mutation of tomato: alterations in the fruit metabolome. New Phytol 166:427–438

Brown SC, Kruppa G, Dasseux JL (2005) Metabolomics applications of FT-ICR mass spectrometry. Mass Spectrom Rev 24:223–31

Carrari F, Fernie AR (2006) Metabolic regulation underlying tomato fruit development. J Exp Bot 57:1883–1897

Damian D, OrešičM, Verheij E et al. (2007) Applications of a new subspace clustering algorithm (COSA) in medical systems biology. Metabolomics 3:69–77

Dettmer K, Aronov PA, Hammock BD (2007) Mass-spectrometry based metabolomics. Mass Spectrom Rev 26:51–78

Dunn WB, Bailey NJC, Johnson HE (2005) Measuring the metabolome: current analytical technologies. Analyst 130:606–625

Fait A, Hanhineva K, Beleggia R et al. (2008) Reconfiguration of the achene and receptacle metabolic networks during strawberry fruit development. Plant Physiol 148:730–750

Fernie AR, Trthewey RN, Krotzky AJ et al. (2004) Metabolite profiling: from diagnostics to systems biology. Nat Rev 5:1–7

Fiehn O, Kopka J, Dörmann P et al. (2000) Metabolite profiling for plant functional genomics. Nat Biotechnol 18:1157–1161

Fiehn O, Sumner LW, Rhee SY et al. (2007) Minimum reporting standards for plant biology context information in metabolomic studies. Metabolomics 3:195–201

Fraser PD, Bramley PM (2006) Metabolic profiling and quantification of carotenoids and related isoprenoids in crop plants. In: Saito K, Dixon R, Willmitzer L (eds) Plant metabolomics, Heidelberg, Springer,Germany, pp 229–240
Fraser PD, Pinto MES, Holloway DE et al. (2000) Application of high-performance liquid chromatography with photodiode array detection to the metabolic profiling of plant isoprenoids. Plant J 24:551–558
Fraser PD, Enfissi EMA, Goodfellow M et al. (2007a) Metabolite profiling of plant carotenoids using matrix assisted laser desorption ionization time of flight mass spectrometry. Plant J 49:552–564
Fraser PD, Enfissi EMA, Halket JM et al. (2007b) Manipulation of phytoene levels in tomato fruit: effects on isoprenoids, plastids, and intermediary metabolism. Plant Cell 19:3194–3211
Fritz JS (2004) Early milestones in the development of ion-exchange chromatography: a personal account. J Chromatogr A 1039:3–12
Fulton TM, Bucheli P, Voirol E et al. (2002) Quantitative trait loci (QTL) affecting sugars, organic acids and other biochemical properties possibly contributing to flavour, identified in four advanced backgross populations of tomato. Euphytica 127:163–177
Gillaspy G, Ben-David H, Gruissem W (1993) Fruits: a developmental perspective. Plant Cell 5:1439–1451
Glinski M, Weckwerth W (2006) The role of mass spectrometry in plant systems biology. Mass Spectrom Rev 25:173–214
Giovannoni JJ (2007) Fruit ripening mutants yield insights into ripening control. Curr Opin Plant Biol 10:283–289
Giovannoni JJ, Noensie EN, Ruezinsky DM et al. (1995) Molecular genetic analysis of the ripening-inhibitor and non-ripening loci of tomato: A first step in genetic map-based cloning of fruit ripening genes. Mol Gen Genet 248:195–206
Halket JM, Waterman D, Przyborowska AM et al. (2005) Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. J Exp Bot 56:219–243
Hall RD (2005) Plant metabolomics: from holistic hope, to hype, to hot topic. New Phytol 169:453–468
Hancock JF (2000) Strawberries. In A Erez (ed) Temperate fruit crops in warm climates. Kluwer Academic, Dordrecht, the Netherlands
Hannum SM (2004) Potential impact of strawberries on human health: a review of the science. Crit Rev Food Sci Nutr 44:1–17
Hukkanen AT, Kokko HI, Buchala AJ et al. (2007) Benzothiadiazole induces the accumulation of phenolics and improves resistance to powdery mildew in strawberries. J Agric Food Chem 55:1862–1870
Iijima Y, Nakamura, Y, Ogata Y et al. (2008) Metabolite annotations based on the integration of mass spectral information. Plant J. 54:949–962
Katajamaa M, Orešič M (2005) Processing methods for differential analysis of LC/MS profile data. BMC Bioinformatics 6:179
Katajamaa M, Orešič M (2007) Data processing for mass spectrometry-based metabolomics. J Chromatogr A 1158:318–328
Ketchum REB, Croteau RB (2006) The taxus metabolome and the elucidation of the Taxol biosynthetic pathway in cell suspension cultures. In: Saito K, Dixon R, Willmitzer L (eds) Plant metabolomics, Springer, Heidelberg, Germany, pp 291–308
Koponen JM, Happonen AM, Mattila PH et al. (2007) Contents of anthocyanins in selected foods consumed in Finland. J Agric Food Chem 55:1612–9
Krishnan P, Kruger NJ, Ratcliffe RG (2005) Metabolite fingerprinting and profiling in plants using NMR. J Exp Bot 56:255–265
Kähkönen MP, Hopia AI, Heinonen M (2001) Berry phenolics and their antioxidant activity. J Agric Food Chem 49:4076–4082
Lange BM (2006) Integrative analysis of metabolic networks: from peaks to flux models? Curr Opin Plant Biol 9:220–226
Lippman ZB, Semel Y, Zamir D (2007) An integrated view of quantitative trait variation using tomato interspecific introgression lines. Curr Opin Genet Dev 17:545–552

Mattoo AK, Sobolev AP, Neelam A et al. (2006) Nuclear magnetic resonance spectroscopy-based metabolite profiling of transgenic tomato fruit engineered to accumulate spermidine and spermine reveals enhanced anabolic and nitrogen-carbon interactions. Plant Phys 142:1759–1770

Mintz-Oron S, Mandel T, Rogachev I et al. (2008) Gene expression and metabolism in tomato fruit surface tissues. Plant Phys. 147:823–851

Moco S, Bino RJ, Vorst O et al. (2006) A liquid chromatography–mass-spectrometry-based metabolome database for tomato. Plant Physiol 141:1205–1218

Moco S, Capanoglu E, Tikunov Y et al. (2007) Tissue specialization at the metabolite level is perceived during the development of tomato fruit. J Exp Bot 58:4131–4146

Moing A, Renauld C, Gaudillere M et al. (2001) Biochemical changes during fruit development of four strawberry cultivars. J Am Soc Hort Sci 126:394–403

Morgenthal K, Weckwerth W, Steuer R (2006) Metabolomic networks in plants: transitions from pattern recognition to biological interpretation. BioSystems 83:108–117

Mullen W, Yokota T, Lean MEJ (2003) Analysis of ellagitannins and conjugates of ellagic acid and quercetin in raspberry fruits by LC-MS^n. Phytochemistry 64:617–624

Neelam A, Cassol T, Mehta RA et al. (2008) A field grown transgenic tomato line expressing higher levels of polyamines reveals legume cover crop mulch-specific perturbations in fruit phenotype at the levels of metabolite profiles, gene expression, and agronomic characteristics. J Exp Bot 59:2337–2346

Puupponen-Pimiä R, Nohynek L, Hartmann-Schmidlin S et al. (2005) Berry phenolics selectively inhibit the growth of intestinal pathogens. J Appl Microbiol 98:991–1000

Robles P, Pelaz S (2005) Flower and fruit development in *Arabidopsis thaliana*. Int J Dev Biol 49:633–643

Roessner U, Luedemann A, Brust D, Fiehn O, Linke T, Willmitzer L, Fernie AR (2001) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. Plant Cell 13:11–29

Roessner-Tunali U, Hegemann B, Lytovchenko A et al. (2003) Metabolic profiling of transgenic tomato plants overexpressing hexokinase reveals that the influence of hexose phophorylation diminishes during fruit development. Plant Physiol 133:84–99

Ryan D, Robards K (2006) Metabolomics: the greatest omics of them all? Anal Chem 78:7954–7958

Saito K, Dixon R, Willmitzer L (2006) (eds) Plant metabolomics. Springer, Heidelberg, Germany

Schauer N, Semel Y, Roessner U et al. (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nat Biotech 24:447–454

Schauer N, Semel Y, Balbo I et al. (2008) Mode of inheritance of primary metabolic traits in tomato. Plant Cell 20:509–523

Seger C, Sturm S (2007) Analytical aspects of plant metabolite profiling platforms: Current standings and future aims. J Proteome Res 6:480–497

Shinbo Y, Nakamura Y, Altaf-Ul-Amin M et al. (2006) KNApSAcK: a comprehensive species-metabolite relationship database. In: Saito K, Dixon RA, Willmitzer L (eds) Biotechnology in agriculture and Forestry, vol. 57. Plant metabolomics Springer, Berlin pp.165–181

Tikunov Y, Lommen A, deVos CHR et al. (2005) A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. Plant Physiol 139:1125–1137

Unger KK (2004) Scientific achievements of Jack Kirkland to the development of HPLC and in particular to HPLC silica packings – a personal perspective. J Chromatogr *A* 1060:1–7

Ward JL, Baker JM, Beale MH (2007) Recent applications of NMR spectroscopy in plant metabolomics. FEBS J 274:1126–1131

Werner RA, Rossmann A, Schwartz C et al. (2004) Biosynthesis of gallic acid in Rhus typhina: discrimination between alternative pathways from natural oxygen isotope abundance. Phytochemistry 65:2809–2813

Villas-Boas SG, Mas S, Åkesson M et al. (2005) Mass spectrometry in metabolome analysis. Mass Spectrom Rev 24:613–646

Chapter 30
Genetic Engineering in Floriculture

Yoshikazu Tanaka and Ryutaro Aida

Abstract Numerous attractive floricultural crops have been developed by extensive hybridization and mutational breeding which suffer from genetic constraint intrinsic to each plant species. Breeding by utilizing genetic engineering has liberated such constraint and any genes from any organisms can be used to make novel floricultural crops. Novel violet/blue colored carnation and rose have been developed by expressing flavonoid biosynthetic genes from heterologous plant species and the carnation has been successfully commercialized. Flowers with modified scents, longer vase life and modified shapes have been also developed. Incorporating progressing plant science will give opportunities to generate novel flowers that appeal consumers.

30.1 Introduction

People have always loved flowers and used flowers to make themselves or their friends happy. At a glance, flowers are innocently lovely; in reality, however, they are bound by modern industrialization and technology (Stewart 2007). The worldwide production value of floricultural crops is estimated to be around 50 billion Euros, and the retail value, 100–150 billion Euros (Chandler and Tanaka 2007). The market consists of an uncountable number of species. Flowers used by florists and in gardens are generally cultivated cultivars that have been made through artificial breeding, including extensive hybridization of an unnatural combination of species and artificial or spontaneous mutations. Recently, genetic engineering has become a method of choice to generate new cultivars.

Y. Tanaka (✉)
Institute for Plant Science, Suntory Holdings Ltd, 1-1-1 Wakayamadai, Shimamoto, Mishima, Osaka 618-8503, Japan
e-mail: Yoshikazu_Tanaka@suntory.co.jp

R. Aida
National Institute of Floricultural Science, Fujimoto 2-1, Tsukuba, Ibaraki 305-8519, Japan

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_30, © Springer Science+Business Media B.V. 2010

Rapid progress of plant sciences has produced (i) useful genes (or molecular tools) to improve plants, (ii) plant transformation systems, and (iii) artificial regulation of trans- or endogenous genes in a transgenic plant. The advantage of plant breeding by genetic engineering (often called molecular breeding) is that plant breeders are not bound by the limits of the genetic diversity of a specific species; with genetic engineering, they can incorporate any genes from any organism to generate a variety with a novel character that is often highly appreciated by consumers.

In this chapter, we will describe what has been achieved in the field of molecular breeding of floricultural crops with a special focus on consumer traits, as called by Chandler and Tanaka (2007), i.e., traits that directly appeal to consumers. Producer traits, such as biotic or abiotic stress resistance, which are mutually important with regard to edible crops, have already been reviewed (Chandler and Tanaka 2007; Tanaka et al. 2005) and will not be discussed here.

30.2 Principles of Molecular Breeding

Several technological steps are necessary for successful molecular breeding. They include the isolation of useful genes, the establishment of a transformation system, and the regulation of the expression of a transgene. In field trials and commercialization, the clearing of complicated regulatory procedures at the international level has been intrinsic to genetically modified (GM) plants. Among these steps, it is noteworthy to stress the importance of plant transformation. Since the performance of a transgene greatly depends on the transgenic lines, many (from dozens to hundreds) transgenic lines containing the same transgene have to be generated to obtain lines exhibiting desirable phenotypes, and thus it is essential to establish an efficient transformation system. However, the plant transformation efficiency depends on the species and cultivars to a large extent. Tedious and laborious efforts for many years are often necessary to optimize transformation protocols. Transformable floricultural crops have been summarized in reviews (Chandler and Tanaka 2007; Tanaka et al. 2005).

The transcription of transgenes can be regulated by using constitutive or spatially or temporally regulated promoters. Downregulation of endogenous genes is often necessary to obtain desirable phenotypes, and it can be efficiently achieved by RNAi suppression. Furthermore, the transcription of the double-strand RNA of a target gene causes such suppression (Waterhouse et al. 1998).

In order to generate a plant with a specific color or scent, a specific compound has to be accumulated by engineering a metabolic pathway (Dixon 2005; Tanaka and Ohmiya 2008). This can be accomplished by the expression of a transgene that redirects the pathway toward the compound. At the same time, the genes of the pathways that compete against the transgene have to be downregulated, or a host that lacks the competing pathway has to be chosen. Substrate availability of an introduced enzyme, degradation/modification or sequestration to undesirable cellular compartment of a product, and possible deleterious effects to plants by the product should also be avoided to achieve successful engineering of a biosynthetic pathway.

30.3 Modification of Flower Color

30.3.1 Biosynthetic Pathways

Flower color comes mainly from flavonoids/anthocyanins, carotenoids, and betalains (see recent reviews: Grotewold 2006; Tanaka et al. 2008). The former two classes of pigments are ubiquitous in seed plants and have been modified by genetic engineering. Anthocyanins, a colored class of flavonoids, confer a wide range of color from orange, red, magenta, violet, and blue. Some flavonoids are pale yellow. Carotenoids are more intensely yellow and often red. Both flavonoid and carotenoid biosynthetic pathways have been well characterized, as shown in Figs. 30.1 and 30.2, and the relevant genes have been isolated.

Plant species often generate limited varieties of flower color; for example, rose, carnation, and chrysanthemum lack violet/blue color, and petunia, pansy, and lisianthus lack orange/bright-red color. This is mainly attributed to their genetically limited ability to synthesize such pigments. As a result, hybridization breeding has been severely limited. However, with molecular breeding, a transgenic plant could utilize a pathway from an unrelated species and, thus, produce a wider range of flower color.

The number of hydroxy groups on the B-ring of anthocyanins is crucial to the color. The number is determined by flavonoid 3'-hydroxylase (F3'H) and flavonoid 3', 5'-hydroxylase (F3'5'H), as shown in Fig. 30.1 and a review (Tanaka 2006). Violet/blue and orange/bright-red flowers tend to have delphinidin- and pelargonidin-based anthocyanins, respectively. Many plant species are unable to accumulate them. Cyanidin-based anthocyanins generally confer red/magenta color. They are most ubiquitous distributed in plants and thus manipulation of cyanidin-based anthocyanins gives less impact to flower color than that of delphinidin- and pelargonidin-based anthocyanins.

30.3.2 Engineering Toward Pelargonidin

Petunia is unable to produce pelargonidin-based anthocyanins because petunia dihydroflavonol 4-reductase (DFR) does not utilize dihydrokaempferol as a substrate (Forkamnn and Ruhnau 1987). This is the reason that petunia lacks orange to brick-red color varieties. DFR of many plant species including maize, gerbera, and rose catalyzes the reduction of dihydrokaempferol. The maize *DFR* gene was introduced to a petunia variety accumulating dihydrokaempferol due to a mutation of the *F3'5H* and *F3'H* genes. The resultant petunia plants accumulated pelargonidin and had a brick-red flower color (Meyer et al. 1987). This was the first GM floricultural crop and the first color-modified plant by molecular breeding. Obtained petunia plants were subjected to field trials. The transgenic petunia plants were not phenotypically stable and gave white and variegated flowers. More than one half of them lost their color during a field trial due to epigenetic variation including methy-

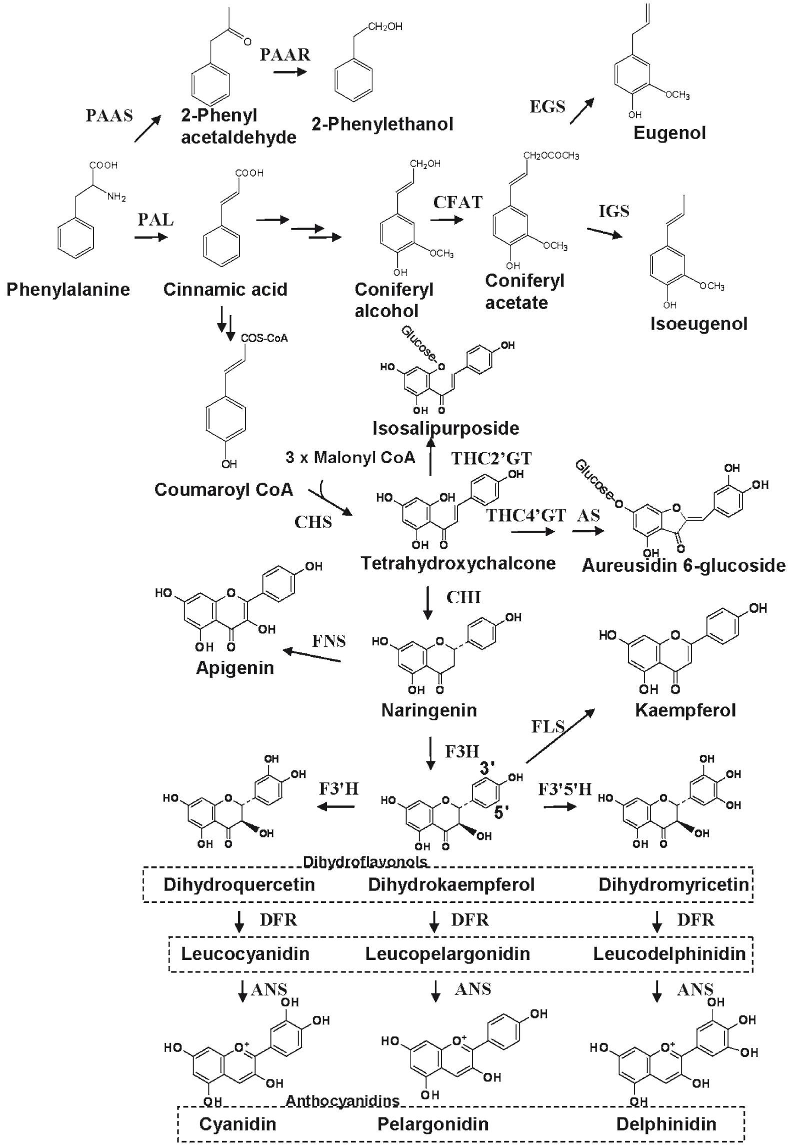


Fig. 30.1 Biosynthetic pathway of major flavonoids/anthocyanidins and some scent benzenoids partly adopted from Pichersky and Dudareva (2007). They are synthesized from phenylalanine. Anthocyanidin is further modified by glycosyltransferase, acyltransferase, and methyltransferse and stored in vacuoles. Abbreviations of flavonoid biosynthetic enzymes: PAL, phenylalanine ammonia lyase; CHS, chalcone synthase; THC2'GT, tetrahydroxychalcone 2'-*O*-glucosyltransferase; CHI, chalcone isomerase; THC4'GT, tetrahydroxychalcone 4'-*O*- glucosyltransferase; AS, aureusidin synthase; F3H, flavanone 3-hydroxylase; F3'H, flavonoid 3'-hydroxylase; F3'5'H, flavonoid 3', 5'-hydroxylase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; FNS, flavone synthase; FLS, flavonol synthase; Glc, glucose. Abbreviations of benzenoid biosynthetic enzymes: PAAS, phenylcetaldehyde synthase; PAAR, phenylacetaldehyde reductase; CFAT, coniferyl alcohol acetyltransferase; EGS, euginol synthase; IGS, isoeuginol synthase

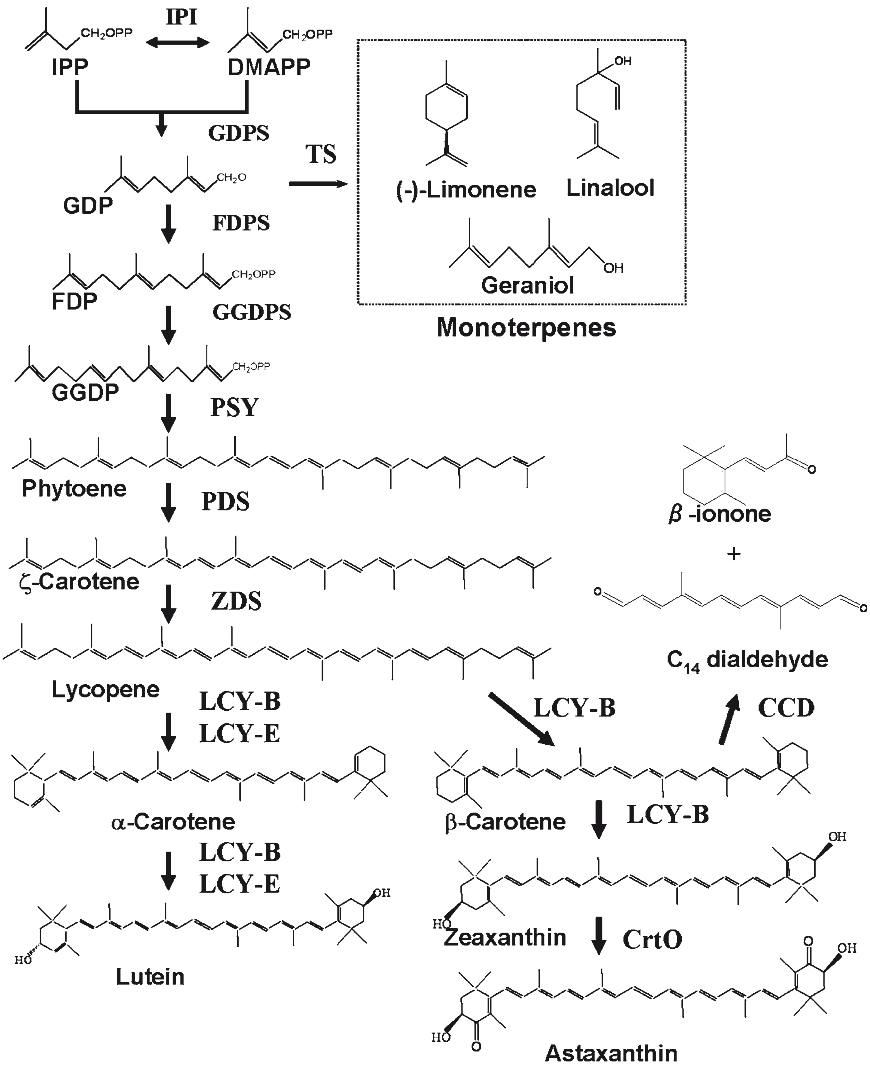


Fig. 30.2 A part of carotenoid and monoterpene biosynthetic pathways. Terpene synthase (TS) is a collective term of enzymes that catalyze synthesis of various terpene. Astaxanthin is not normally produced in plants but have been engineered by the introduction of the *CrtO* gene from algae. Abbreviations include: IPP, isopentenyl pyrophosphate; DMAPP, dimethyl allyl pyrophasphate; GDP, geranyl diphosphate; FDP, farnesyl diphosphate; GGDP, geranyl geranyl diphosphate; GDPS, GDP synthase; FDPS, FDP synthase; GGDPS, GGDP synthase; IPI, isopentenyl phosphate isomerase; PSY, phytoene synthase; PDS, phytoene desaturase; ZDS, ζ-arotene desaturase; LCY-B, lycopene β-cyclase; LCY-E, lycopene ε-cyclase; CCD, carotenoid cleavage dioxygenase

lation (Meyer and Heidmann 1994). A transgenic petunia expressing the maize *DFR* gene was incorporated into a conventional petunia breeding program, and a commercial competent petunia exhibiting strong orange color was successfully achieved (Oud et al. 1995). Although this petunia has not been commercialized, such a combination of molecular and traditional breeding will be a common way to generate novel plants. Similar color petunias accumulating pelargonidin have been obtained from petunia mutants accumulating dihydrokaempferol by constitutively expressing gerbera (Elomaa et al. 1995) or rose (Tanaka et al. 1995) DFR cDNA. Gerbara *DFR* gene performed better than maize *DFR* gene (Elomaa et al. 1995). The selection of the gene source can influence the results even though they encode the same enzymatic activity.

In recent years, more intensive engineering toward pelargonidin has been reported. Downregulation of the *F3'5'H* gene and overexpression of the gerbera *DFR* gene in *Osteospermum hybrida* resulted in pelargonidin accumulation, while overexpression of the gerbera or strawberry *DFR* gene alone did not result in phenotypic changes (Seitz et al. 2007). This is a typical example to show that downregulation of a competing reaction (F3'5'H in this case) is necessary to obtain a desirable phenotype (pelargonidin accumulation in this case). Knockdown of the *F3'H* gene and expression of a rose *DFR* gene in a petunia accumulating cyanidin yielded transgenic orange petunia accumulating pelargonidin. Nevertheless, the amount of flavonols, which are colorless, also increased, indicating that downregulation of flavonol synthase (FLS) was necessary for a more efficient accumulation of pelargonidin (Tsuda et al. 2004). *Nierembergia* only has violet and white varieties. Downregulation of the *F3'5'H* gene yielded white flowers rather than pink flowers, probably due to its strong FLS activity and weak F3'H activity (Ueyama et al. 2006). These results indicated that there is significant competition between colored anthocyanins and colorless flavonols and that downregulation of FLS is necessary to obtain darker flower color in these flowers. Overexpression of the gerbera *DFR* gene and knockdown of the *FLS* and *F3'H* genes resulted in pelargonidin-accumulating tobacco (Nakatsuka et al. 2007). In this report, one chimeric RNAi cassette was used to knock down both genes.

Down regulation of the *F3'5'H* and *F3'H* genes and expression of the rose or pelargonium *DFR* gene in the blue torenia successfully changed the flux from delphinidin to pelargonidin and, thus, yielded red flowers (Chandler and Tanaka 2007; Tanaka and Brugliera 2006). Pelargonidin accumulation by a similar strategy may work to generate red gentian and iris that may command high market value.

30.3.3 Engineering Toward Delphinidin

The *F3'5'H* gene is the key to synthesize delphinidin, and delphinidin should confer violet to blue hues to flowers. The absence of blue/violet varieties in roses, carnations, and chrysanthemums is attributed to the absence of *F3'5'H* gene in these species. The gene was first isolated from petunia (Holton et al. 1993) followed by

many species. Interestingly, expression of the *Campanula medium F3'5'H* gene produced more delphinidin than the petunia and lisianthus genes when expressed in transgenic tobacco (Okinaka et al. 2003), and butterfly pea *F3'5'H* gene produced more delphinidin than the verbena gene in transgenic verbena (Togami et al. 2006). Once again, these results indicate again that the choice of gene source to engineer the pathway is important; however, the reason for any difference in performance is not clear.

Expression of the *F3'5'H* gene in carnation generated delphinidin in the petals and some color shift toward blue. However, the remaining pelargonidin hampers satisfactory color shifts toward blue. Expression of the petunia or pansy *F3'5'H* and the petunia *DFR* (the petunia DFR utilizes dihydromyricetin efficiently but does not utilize dihydrokaempferol) gene in carnation lines deficient in the *DFR* gene successfully redirected the biosynthetic pathway toward delphinidin. The amount of anthocyanins and the flower color were dependent on the constructs, the transgenic events, and the host cultivars. Novel violet-blue transgenic events were subjected to field trials and regulatory procedures on GM. Six kinds of transgenic violet-blue carnations (Florigene Moonseries™) are on the market (Fig. 30.3). The phenotypes are stable after many cycles of vegetative propagation over 8–10 years. This is the first case of commercialization of GM floricultural crops.

Expression of the pansy *F3'5'H* gene in roses yielded delphinidin and color change. The gene was introduced to rose cultivars that have less F3'H activity, large amounts of flavonols (flavonols are colorless but contribute to bluing by their copigment effect), and higher vacuolar pH (anthocyanins tend to be bluer at higher pH) in their petals. A few cultivars produced about 95% delphinidin and a novel blue hue (Katsumoto et al. 2007) (Fig. 30.3). A general release has been granted for two such lines in Japan. More efficient and dominant delphinidin production,



Fig. 30.3 Transgenic carnation and rose accumulating delphinidin by expressing a flavonoid 3', 5'-hydroxylase gene. They wear blue hue that hybridization breeding has not achieved

irrespective of the genetic background of hosts, has been achieved by downregulation of the endogenous *DFR* gene and overexpression of the iris *DFR* gene (*in vivo* functional replacement of DFR) in addition to expression of the pansy *F3'5'H* gene, and a blue hue, which hybridization had not achieved, was thus obtained (Katsumoto et al. 2007).

30.3.4 Engineering Toward Yellow

Some major pot plants, such as pelargonium and begonia, lack yellow cultivars. Chalcones are pale yellow. Tetrahydroxychalcone (THC) is ubiquitously distributed in seed plants but unstable and has to be stabilized by deoxylation or glucosylation to exhibit its color. Expression of *Medicago* (Davies et al. 1998) and the licorice (Tanaka et al. 2005) chalcone reductase gene in white petunia cultivars resulted in very pale yellow color by the accumulation of 6-deoxy chalcones.

More convincing yellow color has been achieved by aurone accumulation. Snapdragons and a few plants contain aurones, a class of flavonoid exhibiting fluorescent yellow color. Expression of snapdragon aureusidin synthase (Nakayama et al. 2000) and THC 4'-glucosyltransferase genes in torenia generated aureusidin glucoside (Fig. 30.1) (Ono et al. 2006). In order to obtain visible yellow color, simultaneous suppression of flavanone 3-hydroxylase (F3H) was necessary. However, expression of the two genes in petunia failed to synthesize aureusidin (unpublished results), and the color in the transgenic torenia is not as yellow as that in the native snapdragon, which may indicate that the two enzymes may be insufficient to synthesize aureusidin in a wide range of plant species.

Carotenoids (Fig. 30.2) confer darker yellow than flavonoids to some flowers, such as rose and chrysanthemum. The ability to synthesize yellow carotenoids in cultivated roses (*Rosa hybrida*) is derived from a Persian yellow wild rose, *R. foetida*. Since carotenoids are essential components of photosystems and carotenoids and phytohormones including gibberellic acid and absidic acid partly share the biosynthetic pathway, engineering a carotenoid pathway may have a detrimental effect and be more challenging than engineering a flavonoid pathway. Owing to their nutrient values, more attention has been given to carotenoid engineering in edible crops than in flowers (Tanaka and Omiya 2008).

In chrysanthemum, yellow flower color is recessive to white: carotenoids are synthesized and, subsequently, degraded by a carotenoid cleavage dioxygenase (CCD) in white color. Suppression of the CCD expression converted the petal color from white to yellow (Ohmiya et al. 2006). The adonis red flower contains red ketocarotenoid, astaxanthin (Cunningham and Gantt 2005). When a gene encoding an algal β-carotene ketolase fused with the tomato phytoene desaturase transit peptide was expressed under the control of a tomato phytoene desaturase promoter in tobacco, astaxanthin and other ketocarotenoids were accumulated to change the color of the nectary tissue from yellow to red (Mann et al. 2000). The petal color of *Lotus japonicus* was successfully changed from light yellow to dark yellow or

orange by expressing a gene encoding a bacterial β-carotene ketorase fused with the transit peptide of a pea RUBISCO small subunit (Suzuki et al. 2007).

30.3.5 Modification of the Amount of Anthocyanins

Downregulation of one biosynthetic gene in flavonoid biosynthesis decreased the amount of anthocyanin and led to white or paler flower color. Since the first report of antisense (van der Krol et al. 1988) or sense (Napoli et al. 1990; van der Krol et al. 1990) suppression of CHS in petunia, similar phenotypic changes have been obtained in rose (Gutterson 1995), carnation (Gutterson 1995; Zuker et al. 2002), chrysanthemum (Courtney-Gutterson et al. 1994), gerbera (Elomaa et al. 1993), lisianthus (Deroles et al. 1998), gentian (Nishihara et al. 2006), and torenia (Aida et al. 2000; Nakamura et al. 2006; Suzuki et al. 2000) by downregulating *CHS, CHI, DFR, F3H* or *ANS* gene. Currently, RNAi suppression is commonly used for downregulation. In addition to attracting pollinators, flavonoids play important roles in stress resistance and they are involved in pollen tube elongation and auxin transport. Therefore, it is not desirable to downregulate flavonoid biosynthesis in the whole plant body, although constitutive promoters have been used in downregulation in the studies mentioned above. From a practical point of view, the application of molecular breeding to generate white cultivars may not be advantageous since white cultivars can be obtained by mutation breeding.

Ectopic accumulation of anthocyanins, such as that in leaves, may have more commercial value. Anthocyanin biosynthesis is regulated by a combination of transcriptional factors (bHLH, Myb, and the WD40 type), as reviewed (Koes et al. 2005), and constitutive expression of the maize *Lc* gene (a bHLH) resulted in colored leaves in tobacco (Lloyd et al. 1992) and petunia (Bradley et al. 1998). The application of this technology to plants used for ornamental foliage may be interesting.

30.4 Modification of Scent

30.4.1 Scent Compounds and Their Biosynthesis

Floral scent is also one of the most important consumer traits for floricultural crops. Unfortunately, most modern cut flowers do not have a strong scent because there is a negative correlation between postharvest vase life and fragrance and fragrance is sacrificed in the course of plant breeding (Chandler and Tanaka 2007; Pichersky and Dudareva 2007). Flowers produce numerous kinds of scent compounds, mainly terpenoids (Fig. 30.2) and phenylpropanoids/benzenoids (Fig. 30.1), which share common pathways with carotenoid and flavonoid biosyntheses, respectively. The structural genes of enzymes that are involved in biosynthesis are being rapidly isolated,

and their application to engineer biosynthesis is also increasing, as reviewed by Aharoni et al. (2006), Schnepp and Dudareva (2006), and Pichersky and Dudareva (2007). Scent compound biosynthesis is mainly regulated at the transcriptional level, and the availability of the substrate of biosynthetic enzymes is often rate-limiting in biosynthesis.

The modification of floral scent by accumulation of a desirable compound is more challenging than flower color modification. Generating transgenic plants expressing these scent genes does not always lead to changes of scent compounds, and successful changes of floral scent detectable by humans are rare. This is attributed to the lack of a proper amount of substrate of introduced enzymes or the modification and subsequent sequestration of products. In addition, floral scent biosynthesis oscillates with the daily cycles, as shown in snapdragon (Dudareva et al. 2000), petunia (Verdonk et al. 2003), and rose (Hendel-Rahmanim et al. 2007; Picone et al. 2004), which may indicate that a transgene should be regulated in a coordinated manner with the endogenous gene of enzymes that synthesize scent compounds or their precursors to obtain desirable phenotypic changes.

30.4.2 Engineering a Terpenoid Biosynthetic Pathway

Terpenoids, the largest class of plant secondary metabolites, are derived from isopentenyl diphosphate (IPP, C5). IPP is supplied by two pathways: the cytosolic mevalonate pathway and the plastid methylerythritol 4-phosphate pathway. In the cytosol, two molecules of IPP and one molecule of its isomer, dimethyl allyl diphosphate (DMADP, C5), are condensed and converted to sesquiterpenes (C15) via farnesyl diphospfate (FDP, C15). In plastids (Fig. 30.2), the condensation of one molecule of IPP and one molecule of DMADP results in geranyl diphosphate (GDP, C10). The addition of one molecule of IPP to GDP yields FDP. Monoterpenes (linalool, limonene, and geraniol) and sesquiterpenes (R-nerolidol and caryophyllene) are synthesized from GDP and FDP, respectively, by the catalysis of various terpene syntheses. These compounds are further modified by oxidation and dehydrogenation.

The first isolated scent gene was the *linalool synthase* (*LIS*) gene from *Clarkia breweri* (Pichersky et al. 1994). In transgenic petunia plants expressing the *C. breweri LIS* gene, linalool was efficiently glucosylated by endogenous glucosyltransferase, and there was no fragrance change (Lucker et al. 2001). Transgenic carnation expressing the same gene emitted only a small amount of linalool and its derivative metabolites (*cis*- and *trans*-linalyl oxides), and no changes in fragrance were detected (Lavy et al. 2002).

Successful modification of fragrance at a detectable level by humans has been achieved. Expression of the *LIS* gene under the control of a tomato late-ripening-specific promoter in tomato fruit that had an abundant supply of GDP yielded linalool and 8-hydroxylinalool in ripening fruits (Lewinsohn et al. 2001). Overexpression of three kinds of lemon monoterpene synthase (*γ-terpinene synthase, limonene synthase,*

and *β-pinene synthase*) genes in tobacco resulted in the emission of β-pinene, limonene, and γ-terpinene and a number of side products in leaves and flowers (Lucker et al. 2004b). Additional introduction of a mint limonene 3-hydroxylase gene resulted in further modification of the terpenoid composition (Lucker et al. 2004a). A high level of terpene production in transgenic tobacco has been accomplished through overexpression of an avian FDP synthase and an appropriate terpene synthase in cytosol or plastids (Wu et al. 2006). Tomato fruit flavor and aroma have been modified by expressing the *Ocimum basilicum geraniol synthase* gene under the control of the tomato ripening-specific polygalacturonase promoter (Davidovich-Rikanati et al. 2007). It would be interesting to express terpenoid scent genes under the control of appropriate promoters in flowers that produce a large amount of carotenoids, such as rose and chrysanthemum, where the supply of substrates should be plenty.

CCD contributes to the formation of aroma as well as pigmentation, as reported above. A petunia and a tomato CCD have been shown to catalyze the cleavage of carotenoids to form β-ionone and geranylacetone, which are important constituents of flavor in petunia and tomato (Simkin et al. 2004a, b).

30.4.3 Benzenoids/Phenylpropanoids

Benzenoids/phenylpropanoids are derived from phenylalanine, and their biosynthetic pathways have been characterized biochemically or molecular-biologically to some extent. RNAi suppression of these genes has been accomplished in petunia. Suppression of benzoic acid/salicylic acid carboxyl methyltransferase resulted in specific depletion of methyl benzoate emission, which was negatively evaluated by human testers (Underwood et al. 2005). Suppression of the phenylacetaldehyde synthase (PAAS, Fig. 30.1) gene eliminated phenylacetaldehyde and 2-phenylethanol (Kaminaga et al. 2006), and that of coniferyl alcohol acetyltransferase (Fig. 30.1) resulted in depletion of the emission of isoeugenol that is synthesized from coniferyl alcohol via an ester of coniferyl alcohol (Dexter et al. 2007; Koeduka et al. 2006). An R2R3-Myb-type transcriptional factor, *ORORANT1*, which is a key regulator of floral scent in petunia, has been isolated. Its suppression diminished benzenoid levels through the decreased synthesis of precursors from the shikimate pathway but did not affect flavonoid bisoynthesis, since flavonoid accumulation precedes scent synthesis (Verdonk et al. 2005).

Suppression of an endogenous gene sometimes yielded unexpected phenotypes. Silencing of benzylalcohol/phenylethanol benzoyltransferase yielded bigger flowers and leaves in petunia in addition to eliminating benzylbenzoate and phenylethylbenzoate as a result of an unknown interaction between the benzenoid pathway and the auxin metabolism (Orlova et al. 2006). Downregulation of F3H in carnation led to an increase of bezenoids as well as paler flower color, indicating the interaction of flavonoid and benzenoid biosynthetic pathways (Zuker et al. 2002).

The successful modification of bezenoid by overexpression of a heterologous gene has not been fully achieved. When the gene of a rose alcohol acetyltransferase

that preferably utilizes geraniol was introduced into petunia, transgenic plants emitted an increased amount of benzyl acetate and phenylethyl acetate. In this case, the substrate availability determined the type of synthesized volatile compounds (Guterman et al. 2006). 2-Phenylethanol is a major constituent of the floral scent of many species and has been shown to be synthesized from phenylalanine via phenethylamine and 2-phenylacetaldehyde in tomato while in petunia one enzyme, PAAS, catalyzes phenylalanine decarboxylation and oxidation to yield 2-phenylacetaldehyde (Kaminaga et al. 2006). The aromatic amino decarboxylase genes have been isolated from tomato. The substrate specificity study of their recombinant enzymes revealed that they utilize tyrosine better than phenylalanine *in vitro*; however, phenylalanine should be the physiological substrate because phenylalanine is more abundant than tyrosine. Overexpression of the genes resulted in increased emissions of 2-phenylacetoaldehyde, 2-phenylethanol, and 1-nitro-2-phenylethane, and the suppression of the genes reduced the emission (Tieman et al. 2006). Tomato phenylacetaldehyde reductase genes have also been isolated. Flowers of transgenic petunia plants expressing the tomato genes emitted higher levels of 2-phenylethanol and lower levels of 2-phenylacetaldehyde (Tieman et al. 2007).

30.5 Improvement of Postharvest Quality

30.5.1 Values of Long Life

Extension of the appreciation period and good postharvest quality are two of the most important characters of floricultural crops and the characters which breeders strive to improve. Such characters are important for both producers and consumers. Preventing flower and/or leaf senescence and obtaining a longer appreciation period are realistic and valuable goals that can be achieved by molecular breeding.

30.5.2 Preventing Flower Senescence

In many floricultural plants, petal senescence is triggered by a phytohormone, ethylene. Transgenic plants with altered ethylene biosynthesis or perception have been reviewed (Chandler and Tanaka 2007; Shibuya and Clark 2006; Stearns and Glick 2003). Ethylene biosynthesis starts with methionine, from which *S*-adenosylmethionine (SAM) is synthesized by the action of SAM synthase. SAM is converted to 1-aminocyclopropane 1-carboxylic acid (ACC) by ACC synthase, and ACC is converted to ethylene by the action of ACC oxidase (Bleecker and Kende 2000). The inhibition of ethylene biosynthesis is one way to delay flower senescence. Suppression of the ethylene biosynthetic gene and the resultant extension of flower longevity have been achieved on carnation (ACC oxidase [Savin et al. 1995], ACC Synthase [Tanaka et al. 2005]), begonia (ACC oxidase

[Hvoslef-Eide et al. 1995]) and torenia (ACC oxidase [Aida et al. 1998]). These plants showed a significant increase in flower longevity due to their decreased ethylene production. However, they are still sensitive to exogenous ethylene, and this may reduce flower longevity.

Conferring plants ethylene insensitivity by inhibition of ethylene perception and/or signal transduction pathway is a more powerful way to delay senescence. Ethylene perception and signal transduction have been well established, as reviewed (Chen et al. 2005), and suppression of ethylene perception or the signal transduction pathway has been reported on several ornamental plants. Expression of a dominant negative mutant ethylene receptor gene, such as *etr1-1* of *Arabidposis*, *boers* of *Brassica*, or *Cm-ETR1/H69A* of melon, has been commonly used to inhibit ethylene perception. By the expression of these mutated ethylene receptor genes, extension of flower longevity has been achieved in carnation (Bovy et al. 1999), petunia (Wilkinson et al. 1997), (Shaw et al. 2002), *Nemesia* (Cui et al. 2004), and *Campanula* (Sriskandarajah et al. 2007). Suppression of a gene, an *EIN2* homologue of petunia, in the ethylene signaling transduction pathway and the resultant extension of flower longevity have also been reported for petunia (Shibuya et al. 2004). The flower life of the transgenic plants reported above was generally about three times longer than that in control plants. Alternatively, in petunia, the senescence-specific expression of the *IPT* gene, a gene encoding isopentenyl transferase, which catalyzes the first and rate-limiting step in the cytokinin biosynthesis pathway, increased flower longevity accompanying a reduction of ethylene sensitivity (Chang et al. 2003).

Unfortunately, none of the transgenic plants has been commercialized. This may be because there are other alternative non-GM ways to delay senescence caused by ethylene, i.e., the current treatment of cut flowers by silver ion (silver nitrate or silver thiosulphate) is highly effective and inexpensive, and the constitutive expression of a mutant receptor gene may be detrimental to the plants, as reviewed by Chandler and Tanaka (2007). Tobacco transformed with the *etr1-1* gene showed decreased disease resistance (Knoester et al. 1998).

A strategy for the prevention of flower senescence on ethylene-sensitive ornamental plants has been established, as reported above; however, the strategy would not be applicable to plants that are less sensitive to ethylene. Petal senescence is related with programmed cell death (Rogers 2006). Future studies are expected reveal the mechanisms of the programmed cell death and to lengthen flower life by preventing the expression of the related genes. The lengthening of post-harvest life by genetic transformation will contribute to increasing the value of ornamental plants.

30.5.3 Preventing Leaf Senescence

Leaf senescence also reduces the value of ornamental plants, as leaves contribute to a plant's ornamental value, as do flowers. Prevention of leaf senescence is one of the targets of the molecular breeding of ornamental plants. Current strategies to prevent leaf senescence are based on phytohormone physiology by either enhancing cytokinin production or blocking ethylene formation or perception (Gan and Amasino 1997).

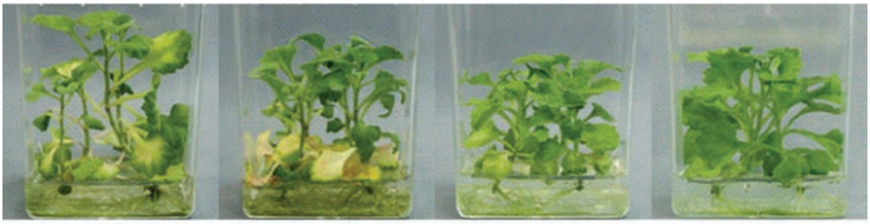


Fig. 30.4 Introduction of a mutated ethylene receptor gene of chrysanthemum produced a transgenic chrysanthemum with reduced ethylene sensitivity and greener leaves for a longer time. *In vitro* plants of the control non-transformed 'Sei-Marine' and the lines of the *mDG-ERS1(etr1-–4)* transformant were treated with ethylene. The transgenic lines were classified into three groups of ethylene sensitivity; the lines with ethylene sensitivity similar to that of 'Sei-Marine' (ES), those with reduced ethylene sensitivity (RES1 lines showing yellowing in one or two of the lowermost leaves), and those with more reduced ethylene sensitivity (RES2 lines showing no leaf yellowing). The photographs show the control 'Sei-Marine' and representatives of the ES, RES1, and RES2 lines, i.e., No. 8, No. 18, and No. 41, respectively, in this order (reprinted from Narumi et al., Copyright (2005), with permission from Elsevier)

The level of endogenous cytokinins that inhibit leaf senescence usually drops with the progression of leaf senescence (Gan and Amasino 1996); thus, overproduction of cytokinins would prevent leaf senescence. The expression of *IPT* will result in the production of cytokinins. A senescence-specific SAG12 promoter was used to direct *IPT* expression in tobacco (Gan and Amasino 1995), *Nicotiana alata* (Schroeder et al. 2001), and petunia (Clark et al. 2004). A cold-inducible promoter, cor15a, was also used to direct *IPT* expression in petunia and chrysanthemum (Khodakovskaya et al. 2005). Those attempts succeeded in sufficiently preventing leaf senescence.

In contrast to cytokinins, ethylene treatment often promotes leaf senescence. Blocking ethylene production or signal transduction would prevent leaf senescence in transgenic plants. The introduction of a mutated ethylene receptor gene of chrysanthemum, *mDG-ERS1(etr1-4)*, into chrysanthemum reduced ethylene sensitivity and resulted in greener leaves for a longer period of time (Narumi et al. 2005) (Fig. 30.4). A transgenic chrysanthemum with reduced ethylene sensitivity in addition to prevention of leaf senescence was also obtained by introduction of the *Ethylene-Insensitive* (*EIN3*) repressor gene (Takako Narumi, personal communication) using the CRES-T method, with which transcriptional factors are converted into dominant-negative (Hiratsu et al. 2003).

30.6 Modification of Plant Shapes

30.6.1 Plant Morphogenesis

Flower and plant shape are among the most important characters of ornamental plants. The creation of novel and attractive shapes will add ornamental value to floricultural crops. Thanks to the rapid understanding of plant morphogenesis in

terms of genetics and molecular biology, the modification of flower and/or plant shape is a promising goal of the molecular breeding of ornamental plants. In addition to the molecular tools described below, the *CYCLOIDEA* gene, which controls floral symmetry in snapdragon (Luo et al. 1996) and its counterparts in legumes (Citerne et al. 2003), have been obtained. Currently, it is possible to modify flower and plant shapes with overexpressing heterologous genes or downregulating endogenous genes; however, novel phenotypes that appeal to consumers have not yet been developed. However, further research and development will lead to the development of plants with flowers of novel shapes that command high market value.

30.6.2 Modification of Flower Shape

Flower shape is one of the most important characters in ornamental plants. The creation of new flower shapes has been a major breeding target. In general, a double flower type has higher commercial value than a single one. Double flowers are caused by the conversion of sepals and/or stamens and/or carpels to petals, increasing floral organ number, floral indeterminancy, or, secondly flower production (Vishnevetsky and Meyerowitz 2002). The molecular mechanisms of the formation of doubleness in flowers have been clarified to some extent. Key transcriptional factors for the identification of floral organs have been clarified by analyzing model plants, such as *Arabidopsis*. The ABC model (Coen and Meyerowitz 1991) and its modified version (Theißen 2001) are known to be applicable to a broad range of plants (Kim et al. 2005). The ABC model proposed that three functionally different genes, i.e., A, B, and C, specified the four-whorl structure of the flower. Gene A is responsible for sepal development in the first (outermost) whorl. Both A and B together specify the petals in the second whorl. B and C determine the stamens in the third whorl, and C alone specifies the carpels in the fourth whorl (Coen and Meyerowitz 1991).

Suppression of the C gene would produce a double-flower phenotype (no stamens and no carpels) with loss of floral determinacy, as reported on *Arabidopsis* C gene *AGAMOUS* (Bowman et al. 1989). In the Japanese morning glory, insertion of a transposable element into a C function gene *DUPLICATED* causes a double flower (Nitasaka 2003). In gerbera, the ray florets in transformants with antisense gerbera-*AGAMOUS* formed corolla-like organs in the third whorl, and all floret types formed carpelloid- and pappus-like organs in the fourth whorl; however, they maintained floral determinacy (Yu et al. 1999). The flower shape of chrysanthemum was also modified by suppressing the chrysanthemum-*AGAMOUS* gene with an antisense transgene (Aida et al. 2008). The pistil of each ray floret was transformed to several corolla-like tissues (secondary corolla) and a pistil-like tissue (Fig. 30.5). The structure of the surface cells of filaments transformed into corolla-like cells.

The other strategy for the development of a double flower is over-expression of the B gene for converting sepals to petals in whorl 1. Ectopic expression of Antirrhinum majus B genes *DEF* and *GLO* in transgenic tobacco resulted in the

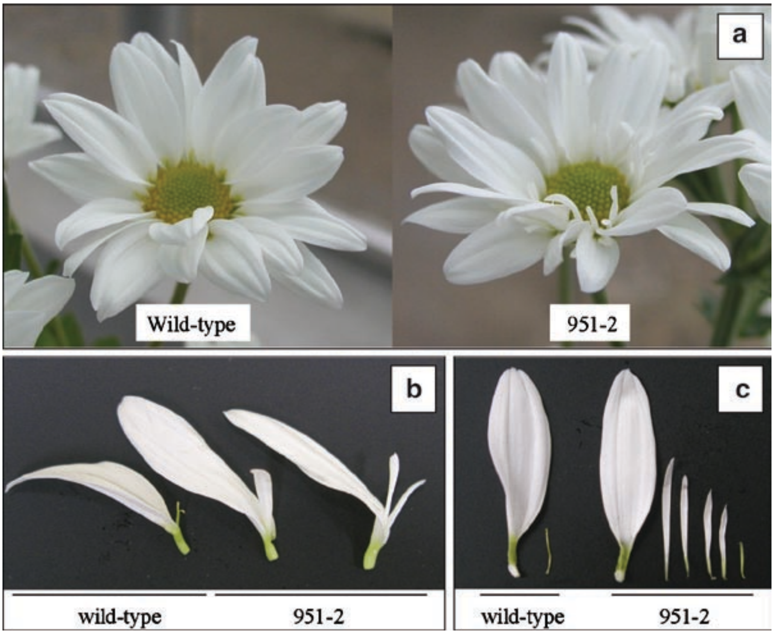


Fig. 30.5 Modified ray florets of a transgenic chrysanthemum with reduced chrysanthemum-*AGAMOUS* gene expression. The pistil of each ray floret was changed to several corolla-like tissues (secondary corolla) and a pistil-like tissue. (**a**) Capitulum, (**b**) intact ray florets, and (**c**) ray florets divided into each tissue type (reprinted from Aida et al., Copyright (2008), with permission from Japanese Society for Plant Cell and Molecular Biology)

conversion of sepals to petals (Davies et al. 1996). An almost identical result was obtained on *Torenia fournieri* by over-expressing the *DEF* and *GLO* genes (Takashi Handa, personal communication).

Application of a synthetic cytokinin, forchlorfenuron (CPPU), to inflorescences of torenia induced flower shape modification, such as serrate petals, incised petals, a paracorolla, and an increased number of floral organs (Nishijima and Shima 2006). These morphological changes occurred systematically depending on the floral stage at the time of CPPU application. The introduction of cytokinin-related genes seemed to change flower shapes on some ornamental plants, such as torenia.

30.6.3 Modification of Plant Form

Plant form determines the collective impression of ornamental plants. The creation of new plant forms can lead to new cultivars with new ornamental values. For example, a dwarf phenotype is suitable for potted plants, and breeding of such a phenotype would contribute to a new use for the plants.

Plant form can be changed by transgenically controlling hormonal regulation. Transformation with *rol* genes from the bacterium *Agrobacterium rhizogenes* is

known to affect cytokinin and auxin concentration in infected plants and is useful to produce dwarf and bushy plants (Casanova et al. 2005). Introduction of *rol* genes has been reported on many floricultural crops, and this method has proven to be useful for producing dwarf and bushy phenotypes. Currently reported rol gene-transformed ornamental plants are *Angelonia salicariifolia, Antirrhinum majus, Begonia tuberhybrida, Catharanthus roseus*, carnation, chrysanthemum, *Datura arborea, Datura sanguinea, Eustoma grandiflorum, Gentiana* sp., *Ipomoea trichocarpa, Lilium longiflorum, Limonium* sp., *Nierembergia scoparia, Osteospermum ecklonis*, *Pelargonium* sp., petunia, rose, *Rudbeckia hirta*, and *Salpiglossis sinuate*, as reviewed by Casanova et al. (2005). It is obvious that introducing *rol* genes is useful for yielding modified plant architecture for ornamental species. *RolC*-transformed plants might increase their number of flowers or/and might show early flowering.

Another method to modify plant height is to use the mutated *GAI* (gibberellic acid insensitive) gene. By introduction of the mutated *GAI* gene, dwarf plants have been obtained on chrysanthemum (Petty et al. 2005) and petunia (Tanaka et al. 2005). Regulation of a gene of the cytokinin biosynthetic pathway is also useful for modifying plant shape. Transgenic *Nicotiana* transformed with an *IPT* gene under the control of an *Arabidopsis* leaf senescence promoter showed dwarf and bushy phenotypes accompanying the prevention of leaf senescence (Schroeder et al. 2001).

The *CENTRORADIALIS* (*CEN*) gene of snapdragon is necessary for the indeterminate growth of the shoot meristem in snapdragon. Its overexpression in tobacco (a determinate species) extended the vegetative phase with more leaves and a taller architecture and delayed flowering by more than 10 months (Amaya et al. 1999). This gene may be useful to delay flowering and generate taller plants.

The overexpression of a petunia zinc-finger type transcription factor, Lateral-shoot Inducing Factor (LIF), in petunia showed a dramatic increase in the number of lateral shoots (Nakagawa et al. 2005). In chrysanthemum, several other genes have been shown to affect plant architecture; the *phytochrome B1* gene made slightly dwarf plants (Zheng et al. 2001), the *Ls*-like gene made a reduced-branching plant (Han et al. 2007), and the *IbMADS4* gene made dwarf and bushy plants (Aswath et al. 2004).

There are several methods to modify plant architecture for enhancing ornamental values. Transgenic ornamental plants with novel and attractive forms that fit market needs are anticipated for future commercialization.

30.7 Control of Flowering – Florigen

In a Japanese fairy tale, as soon as a good old man scattered magic ash among cherry trees, they started to flower. Such ash is the most desirable in the floricultural industry. Florigen that may corresponds to the magic ash was proposed as a kind of phytohormone that promote flowering about 70 years ago, but florigen was not successfully isolated, and even its existence was questioned. It was not until recently that proteins encoded by the *FT* (*flowering locus T*) gene in *Arabidopsis* and its ortholog Hd3a protein in rice meet the definition of florigen.

Loss of *FT* caused delay in flowering, and overexpression of the *FT* gene promoted flowering independently of the photoperiod in *Arabidopsis* (Kardailsky et al. 1999; Kobayashi et al. 1999). Similarly, overexpression of the *FT* gene (Seiichi Fukai, personal communication) or the *Hd3a* gene (Hiroshi Asao, personal communication) in chrysanthemum resulted in early flowering. Some of the transgenic plants flower during tissue culture in the process of transformation. Trangenic trifoliate orange overexpressing Citrus *FT* cDNA starts flowering as early as 12 weeks after transfer to a greenhouse, while the host plants take several years to do so (Endo et al. 2005). Another protein called FD expressed in the shoot apex is required for *FT* to promote flowering.

FD and FT are interdependent partners through protein interaction and act at the shoot apex to promote floral transition (Abe et al. 2005; Wigge et al. 2005). It has been shown that proteins of FT and its rice homologue of FT (Hd3a) move from the leaf to the apical meristem and induces flowering in rice, which meets the definition of florigen (Corbesier et al. 2007; Tamaki et al. 2007). Ideally, chemicals inducing flowering are desirable.

30.8 Present and Future of Transgenic Flowers

More fundamental progress in plant science is obviously essential to develop attractive transgenic floricultural crops. Since plant metabolic pathways and hormonal regulation extensively cross-talk with each other, the modification of characters described here without side effects is more challenging than the simple expression of herbicides or insect-resistant genes that have been successfully achieved in soybean, maize, cotton and canola. More sophisticated regulation of transgene and endogenous genes in transgenic floricultural crops is expected to yield novel flowers that will satisfy consumers.

The value of sales per cultivar in floriculture is not big enough to justify the cost of research and development and the additional cost of clearing regulatory procedures on GM crops (Chandler and Tanaka 2007). Significant relaxation of the regulations is necessary to boost the application and commercialization of genetic engineering of floricultural crops.

References

Abe M, Kobayashi Y, Yamamoto S et al (2005) FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science 309:1052–1056

Aharoni A, Jongsma MA, Kim T-Y et al (2006) Metabolic engineering of terpenoid biosynthesis in plants. Phyochem Rev 5:49–58

Aida R, Kishimoto S, Tanaka Y et al (2000) Modification of flower color in torenia (*Torenia fournieri* Lind.) by genetic transformation. Plant Sci 153:33–42

Aida R, Komano M, Saito M et al (2008) Chrysanthemum flower shape modification by suppression of chrysanthemum-*AGAMOUS* gene. Plant Biotechnol 25:55–59

Aida R, Yoshida T, Ichimura K et al (1998) Extension of flower longevity in transgenic torenia plants incorporating ACC oxidase transgene. Plant Sci 138:91–101

Amaya I, Ratcliffe OJ, Bradley DJ (1999) Expression of *CENTRORADIALIS (CEN)* and *CEN*-like genes in tobacco reveals a conserved mechanism controlling phase change in diverse species. Plant Cell 11:1405–1418

Aswath CR, Mo SY, Kim SH et al (2004) *IbMADS4* regulates the vegetative shoot development in transgenic chrysanthemum (*Dendrathema grandiflora* (Ramat.) Kitamura) Plant Sci 166:847–854

Bleecker AB, Kende H (2000) Ethylene: a gaseous signal molecule in plants. Ann Rev Cell Dev Biol 16:1–18

Bovy AG, Angenent GC, Dons HJM et al (1999) Heterologous expression of the *Arabidopsis etr1-1* allele inhibits the senescence of carnation flowers. Mol Breed 5:301–308

Bowman JL, Smyth DR, Meyerowitz EM (1989) Genes directing flower development in *Arabidopsis*. Plant Cell 1:37–52

Bradley JM, Davies KM, Deroles SC et al (1998) The maize *Lc* regulatory gene up-regulates the flavonoid biosynthetic pathway of Petunia. Plant J 13:381–392

Casanova E, Trillas MI, Moysset L et al (2005) Influence of *rol* genes in floriculture. Biotechnol Adv 23:3–39

Chandler S, Tanaka Y (2007) Genetic modification in floriculture. CRC Crit Rev Plant Sci 26:169–197

Chang H, Jones ML, Banowetz GM et al (2003) Overproduction of cytokinins in petunia flowers transformed with P(SAG12)-IPT delays corolla senescence and decreases sensitivity to ethylene. Plant Physiol 132:2174–2183

Chen YF, Etheridge N, Schaller GE (2005) Ethylene signal transduction. Ann Bot 95:901–915

Citerne HL, Luo D, Pennington RT et al (2003) A phylogenomic investigation of *CYCLOIDEA*-like *TCP* genes in the Leguminosae. Plant Physiol 131:1042–1053

Clark DG, Dervinis C, Barrett JE et al (2004) Drought-induced leaf senescence and horticultural performance of transgenic PSAG12-*IPT* petunias. J Am Soc Hort Sci 129:93–99

Coen ES, Meyerowitz EM (1991) The war of the whorls: genetic interactions controlling flower development. Nature 353:31–37

Corbesier L, Vincent C, Jang S et al (2007) FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. Science 316:1030–1033

Courtney-Gutterson N, Napoli C, Lemieux C et al (1994) Modification of flower color in Florist's Chrysanthemum: production of a white-flowering variety through molecular genetics. Biotechnol 12:268–271

Cui ML, Takada K, Ma B et al (2004) Overexpression of a mutated melon ethylene receptor gene *Cm-ETR1/H69A* confers reduced ethylene sensitivity in a heterologous plant, *Nemesia strumosa*. Plant Sci 167:253–258

Cunningham FXJ, Gantt E (2005) A study in scarlet: enzymes of ketocarotenoid biosynthesis in the flowers of *Adonis aestivalis*. Plant J 41:478–492

Davidovich-Rikanati R, Sitrit Y, Tadmor Y et al (2007) Enrichment of tomato flavor by diversion of the early plastidial terpenoid pathway. Nat Biotechnol 25:899–901

Davies B, Rosa AD, Eneva T et al (1996) Alteration of tobacco floral organ identity by expression of combinations of *Antirrhinum* MADS-box genes. Plant J 10:663–677

Davies KM, Bloor SJ, Spiller GB et al (1998) Production of yellow colour in flowers: redirection of flavonoid biosynthesis in *Petunia*. Plant J 13:259–266

Deroles S, Bradley JM, Schwinn KE et al (1998) An antisense chalcone synthase cDNA leads to novel colour patterns in lisianthus (*Eustoma grandiflorum*) flowers. Mol Breed 4:59–66

Dexter R, Qualley A, Kish CM et al (2007) Characterizatio of a petunia acetyltransferase involved in the biosynthesis of the floral volatile isoeugenol. Plant J 49:265–275

Dixon RA (2005) Engineering of plant natural product pathways. Curr Opin Plant Biol 8:329–336

Dudareva N, Murfitt LM, Mann CJ et al (2000) Developmental regulation of methyl benzoate biosynthesis and emission in snapdragon flowers. Plant Cell 12:949–961

Elomaa P, Helariutta Y, Griesbach RJ et al (1995) Transgenic inactivation in *Petunia hybrida* is influenced by the properties of the foreign gene. Mol Gen Genet 248:645–649

Elomaa P, Honkanen J, Puska R et al. (1993) *Agrobacterium*-mediated transfer of antisense chalcone synthase cDNA to *Gerbera hybrida* inhibits flower pigmentation. Biotechnol 11:508–511
Endo T, Shimada T, Fujii H et al (2005) Ectopic expression of an FT homolog from citrus confers an early flowering phenotype on trifoliate orange (*Poncirus trifoliata* L. Raf.). Transgenic Res 14:703–712
Forkamnn G, Ruhnau B (1987) Distinct substrate specificity of dihydroflavonol 4-reductase from flowers of *Petunia hybrida*. Z. Naturforschung 42c:1146–1148
Gan S, Amasino RM (1995) Inhibition of leaf senescence by autoregulated production of cytokinin. Science 270:1986–1988
Gan S, Amasino RM (1996) Cytokinins in plant senescence: from spray and pray to clone and play. BioEssays 18:557–565
Gan S, Amasino RM (1997) Making sense of senescence. Molecular genetic regulation and manipulation of leaf senescence. Plant Physiol 113:313–319
Grotewold E (2006) The genetics and biochemistry of floral pigments. Annu Rev Plant Biol 57:761–780
Guterman I, Masci T, Chen X et al (2006) Generation of phenylpropanoid pathway-derived volatiles in transgenic plants: rose alcohol acetyltransferase produces phenylethyl acetate and benzyl acetate in petunia flowers. Plant Mol Biol 60:555–563
Gutterson N (1995) Anthocyanin biosynthetic genes and their application to flower color modification through sense suppression. Hort Sci 30:964–966
Han BH, Suh EJ, Lee SY et al (2007) Selection of non-branching lines induced by introducing Ls-like cDNA into chrysanthemum (*Dendranthema grandiflorum* (Ramat.) Kitamura) "Shuho-no-chikara". Sci Hortic 115:70–75
Hendel-Rahmanim K, Masci T, Vainstein A et al (2007) Diurnal regulation of scent emission in rose flowers. Planta 226:1491–1499
Hiratsu K, Matsui K, Koyama T et al (2003) Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in *Arabidopsis*. Plant J 34:733–739
Holton TA, Brugliera F, Lester DR et al (1993) Clonig and expression of cytochrome P450 genes controlling flower colour. Nature 366:276–279
Hvoslef-Eide AK, Fjeld T, Einset JW (1995) Breeding Christmas begonia (*Begonia x cheimantha* Everett) for increased keeping quality by traditional and biotechnological methods. Acta Hortic 405:197–204
Kaminaga Y, Schnepp J, Peel G et al (2006) Plant phenylacetaldehyde synthase is a bifunctional homotetrameric enzyme that catalyzes phenylalanine decarboxylation and oxidation. J Biol Chem 281:23357–23366
Kardailsky I, Shukla VK, Ahn JH et al (1999) Activation tagging of the floral inducer FT. Science 286:1962–1965
Katsumoto Y, Mizutani M, Fukui Y et al (2007) Engineering of the rose flavonoid biosynthetic pathway successfully generated blue-hued flowers accumulating delphinidin. Plant Cell Physiol 48:1589–1600
Khodakovskaya M, Li Y, Li J et al (2005) Effects of *cor15a-IPT* gene expression on leaf senescence in transgenic *Petunia x hybrida* and *Dendranthema x grandiflorum*. J Exp Bot 56:1165–1175
Kim S, Koh J, Yoo MJ et al (2005) Expression of floral MADS-box genes in basal angiosperms: implications for the evolution of floral regulators. Plant J 43:724–744
Knoester M, van Loon L, van Heuvel J et al (1998) Ethylene-insensitive tobacco lacks non-host resistance against soil-borne fungi. Proc Natl Acad Sci USA 95:1933–1937
Kobayashi Y, Kaya H, Goto K et al (1999) A pair of related genes with antagonistic roles in mediating flowering signals. Science 286:1960–1962
Koeduka T, Fridman E, Gang DR et al (2006) Eugenol and isoeugenol, characteristic aromatic constituents of spices, are biosynthesized via reduction of a coniferyl alcohol ester. Proc Natl Acad Sci USA 103:10128–10133
Koes R, Verweij W, Quattrocchio F (2005) Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. Trends Plant Sci 10:236–242
Lavy M, Zuker A, Lewinsohn E et al (2002) Linalool and linalool oxide production in transgenic carnation flowers expressing the *Clarkia breweri* linalool synthase gene. Mol Breed 9:103–111

Lewinsohn E, Schalechet F, Wilkinson J et al (2001) Enhanced levels of the aroma and flavor compound *S*-linalool by metabolic engineering of the terpenoid pathway in tomato fruits. Plant Physiol 127:1256–1265

Lloyd AM, Walbot V, Davis RW (1992) *Arabidopsis* and *Nicotiana* anthocyanin production activated by maize regulators R and C1. Science 258:1773–1775

Lucker J, Bouwmeester HJ, Schwab W et al (2001) Expression of Clarkia *S*-linalool synthase in transgenic petunia plants results in the accumulation of *S*-linalyl-beta-D-glucopyranoside. Plant J 27:315–324

Lucker J, Schwab W, Franssen MC et al (2004a) Metabolic engineering of monoterpene biosynthesis: two-step production of (+)-trans-isopiperitenol by tobacco. Plant J 39:135–145

Lucker J, Schwab W, van Hautum B et al (2004b) Increased and altered fragrance of tobacco plants after metabolic engineering using three monoterpene synthases from lemon. Plant Physiol 134:510–519

Luo D, Carpenter R, Vincent C et al (1996) Origin of floral asymmetry in *Antirrhinum*. Nature 383:794–799

Mann V, Harker M, Pecker I et al (2000) Metabolic engineering of astaxanthin production in tobacco flowers. Nat Biotechnol 18:888–892

Meyer P, Heidmann I (1994) Epigenetic variants of a transgenic petunia line show hypermethylation in transgene DNA: an indication for specific recognition of foreign DNA in transgenic plants. Mol Gen Genet 243:390–399

Meyer P, Heidemann I, Forkmann G et al (1987) A new petunia flower colour generated by transformation of a mutant with a maize gene. Nature 330:677–678

Nakagawa H, Jiang CJ, Sakakibara H et al (2005) Overexpression of a petunia zinc-finger gene alters cytokinin metabolism and plant forms. Plant J 41:512–523

Nakamura N, Fukuchi-Mizutani M, Suzuki K et al (2006) RNAi suppression of the anthocyanidin synthase gene in *Torenia hybrida* yields white flowers with higher frequency and better stability than antisense and sense suppression. Plant Biotechnol 23:13–18

Nakatsuka T, Abe Y, Kakizaki Y et al (2007) Production of red-flowered plants by genetic engineering of multiple flavonoid biosynthetic genes. Plant Cell Rep 26:1951–1959

Nakayama T, Yonekura-Sakakibara K, Sato T et al (2000) Aureusidin synthase: a polyphenol oxidase homolog responsible for flower coloration. Science 290:1163–1166

Napoli CL, Lemieux C, Jorgensen R (1990) Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes *in trans*. Plant Cell 2:279–289

Narumi T, Aida R, Ohmiya A et al (2005) Transformation of chrysanthemum with mutated ethylene receptor genes: *mDG-ERS1* transgenes conferring reduced ethylene sensitivity and characterization of the transformants. Postharvest Biol Technol 37:101–110

Nishihara M, Nakatsuka T, Hosokawa K et al (2006) Dominant inheritance of white-flowered and herbicide-resistant traits in transgenic gentian plants. Plant Biotechnol 23:25–31

Nishijima T, Shima K (2006) Change in flower morphology of *Torenia fournieri* Lind. induced by forchlorfenuron application. Sci Hortic 109:254–261

Nitasaka E (2003) Insertion of an *En/Spm*-related transposable element into a floral homeotic gene *DUPLICATED* causes a double flower phenotype in the Japanese morning glory. Plant J 35:522–531

Ohmiya A, Kishimoto S, Aida R et al (2006) Carotenoid cleavage dioxygenase (CmCCD4a) contributes to white color formation in chrysanthemum petals. Plant Physiol 142:1193–1201

Okinaka Y, Shimada Y, Nakano-Shimada R et al (2003) Selective accumulation of delphinidin derivatives in tobacco using a putative flavonoid 3',5'-hydroxylase cDNA from *Campanula medium*. Biosci Biotechnol Biochem 67:161–165

Ono E, Fukuchi-Mizutani M, Nakamura N et al (2006) Yellow flowers generated by expression of the aurone biosynthetic pathway. Proc Natl Acad Sci USA 103:11075–11080

Orlova I, Marshall-Colon A, Schnepp J et al (2006) Reduction of benzenoid synthesis in petunia flowers reveals multiple pathways to benzoic acid and enhancement in auxin transport. Plant Cell 18:3458–3475

Oud JSN, Schneiders H, Kool AJ et al (1995) Breeding of transgenic orange *Petunia hybrida* varieties. Euphytica 85:403–409

Petty LM, Harberd NP, Carre IA et al (2005) Expression of the *Arabidopsis gai* gene under its own promoter causes a reduction in plant height in chrysanthemum by attenuation of the gibberellin response. Plant Sci 164:175–182

Pichersky E, Dudareva N (2007) Scent engineering: toward the goal of controlling how flowers smell. Trends Biotechnol 25:105–110

Pichersky E, Raguso RA, Lewinsohn E et al (1994) Floral scent production in *Clarkia* (Onagraceae) (I. Localization and developmental modulation of monoterpene emission and linalool synthase activity). Plant Physiol 106:1533–1540

Picone MJ, Clery RA, Watanane N et al (2004) Rhythmic emission of floral volatiles from *Rosa damascena semperflorens* cv. 'Quatre Saisons'. Planta 219:468–478

Rogers HJ (2006) Programmed cell death in floral organs: how and why do flowers die? Ann Bot 97:309–315

Savin KW, Baudinette SC, Graham MW et al (1995) Antisense ACC oxidase RNA delays carnation petal senescence. Hort Sci 30:970–972

Schnepp J, Dudareva N (2006) Floral scent: biosynthesis, regulation and genetic modification. In: Ainsworth C (ed) Flowering and its manipulation. Blackwell

Schroeder KR, Stimart DP, Nordheim EV (2001) Response of *Nicotiana alata* to insertion of an autoregulated senescence-inhibition gene. J Am Soc Hortic Sci 125:523–530

Seitz C, Vitten M, Steinbach P et al (2007) Redirection of anthocyanin synthesis in *Osteospermum hybrida* by a two-enzyme manipulation strategy. Phytochemistry 68:824–833

Shaw JF, Chen HH, Tsai MF et al (2002) Extended flower longevity of *Petunia hybrida* plants transformed with boers, a mutated *ERS* gene of *Brassica oleracea*. Mol Breed 9:211–216

Shibuya K, Clark DG (2006) Ethylene; current status and future directions of using transgenic techniques to improve flower longevity of ornamental crops. J Crop Improv 18:391– 412

Shibuya K, Barry KG, Ciardi JA et al (2004) The central role of *PhEIN2* in ethylene responses throughout plant development in petunia. Plant Physiol 136:2900–2912

Simkin AJ, Schwartz SH, Auldridge M et al (2004a) The tomato carotenoid cleavage dioxygenase 1 genes contribute to the formation of the flavor volatiles beta-ionone, pseudoionone, and geranylacetone. Plant J 40:882–892

Simkin AJ, Underwood BA, Auldridge M et al (2004b) Circadian regulation of the PhCCD1 carotenoid cleavage dioxygenase controls emission of beta-ionone, a fragrance volatile of petunia flowers. Plant Physiol 136:3504–3514

Sriskandarajah S, Mibus H, Serek M (2007) Transgenic *Campanula carpatica* plants with reduced ethylene sensitivity. Plant Cell Rep 26:805–813

Stearns JC, Glick BR (2003) Transgenic plants with altered ethylene biosynthesis or perception. Biotechnol Adv 21:193–210

Stewart A (2007) Flower confidential. Algonquin books of Chapel Hill, Chapel Hill, NC.

Suzuki K, Zue H, Tanaka Y et al (2000) Flower color modifications of *Torenia hybrida* by cosuppression of anthocyanin biosynthesis genes. Mol Breed 6:239–246

Suzuki S, Nishihara M, Nakatsuka T et al (2007) Flower color alteration in *Lotus japonicus* by modification of the carotenoid biosynthetic pathway. Plant Cell Rep 26:951–959

Tamaki S, Matsuo S, Wong HL et al (2007) Hd3a protein is a mobile flowering signal in rice. Science 316:1033–1036

Tanaka Y (2006) Flower colour and cytochromes P450. Phyochem Rev 5:283–291

Tanaka Y, Brugliera F (2006) Flower colour. In: Ainsworth C (ed) Flowering and its manipulation. Blackwell.

Tanaka Y, Ohmiya A (2008) Seeing is believing. Engineering a plant pigment biosynthetic pathway. Curr Opin Biotechnol 19:190–197

Tanaka Y, Fukui Y, Fukuchi-Mizutani M et al (1995) Molecular cloning and characterization of *Rosa hybrida* dihydroflavonol 4-reductase. Plant Cell Physiol 36:1023–1031

Tanaka Y, Katsumoto Y, Brugliera F et al (2005) Genetic engineering in floriculture. Plant Cell, Tissue Organ Cult 80:1–24

Tanaka Y, Sasaki N, Ohmiya A (2008) Plant pigments for coloration. Plant J 54:733–749

Theißen G (2001) Development of floral organ identity: stories from the MADS house. Curr Opin Plant Biol 4:75–85

Tieman D, Taylor M, Schauer N et al (2006) Tomato aromatic amino acid decarboxylases participate in synthesis of the flavor volatiles 2-phenylethanol and 2-phenylacetaldehyde. Proc Natl Acad Sci USA 103:8287–8292

Tieman DM, Loucas HM, Kim JY et al (2007) Tomato phenylacetaldehyde reductases catalyze the last step in the synthesis of the aroma volatile 2-phenylethanol. Phytochemistry 68:2660–2669

Togami J, Tamura M, Ishiguro K et al (2006) Molecular characterization of the flavonoid biosynthesis of *Verbena hybrida* and the functional analysis of verbena and *Clitoria ternatea F3'5'H* genes in transgenic verbena. Plant Biotehcnol 23:5–11

Tsuda S, Fukui Y, Nakamra N et al (2004) Flower color modification of *Petunia hybrida* commercial varieties by metabolic engineering. Plant Biotechnol 21:377–386

Ueyama Y, Katsumoto Y, Fukui Y et al (2006) Molecular characterization of the flavonoid biosynthetic pathway and flower color modification of *Nierembergia* sp. Plant Biotechnol 23:19–24

Underwood BA, Tieman DM, Shibuya K et al (2005) Ethylene-regulated floral volatile synthesis in petunia corollas. Plant Physiol 138: 255–266

van der Krol AR, Lenting PE, Veenstra J et al (1988) An antisense chalcone synthase gene in transgenic plants inhibits flower pigmentation. Nature 333:866–869

van der Krol AR, Mur LA, Beld M et al (1990) Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. Plant Cell 2:291–299

Verdonk JC, Haring MA, van Tunen AJ et al (2005) *ODORANT1* regulates fragrance biosynthesis in petunia flowers. Plant Cell 17:1612–1624

Verdonk JC, Ric de Vos CH, Verhoeven HA et al (2003) Regulation of floral scent production in petunia revealed by targeted metabolomics. Phytochemistry 62:997–1008

Vishnevetsky M, Meyerowitz EM (2002) Molecular control of flower development. In: Vainstein A (ed) Breeding for ornamentals: classical and molecular approaches. Kluwer, Norwell, MA.

Waterhouse PM, Graham MW, Wang M-B (1998) Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. Proc Natl Acad Sci USA 95:13959–13964

Wigge PA, Kim MC, Jaeger KE et al (2005) Integration of spatial and temporal information during floral induction in *Arabidopsis*. Science 309:1056–1059

Wilkinson JQ, Lanahan MB, Clark DG et al (1997) A dominant mutant receptor from *Arabidopsis* confers ethylene insensitivity in heterologous plants. Nat Biotechnol 15:444–447

Wu S, Schalk M, Clark A et al (2006) Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants. Nat Biotechnol 24:1441–1447

Yu D, Kotilainen M, Pollanen E et al (1999) Organ identity genes and modified patterns of flower development in *Gerbera hybrida* (Asteraceae). Plant J 17:51–62

Zheng ZL, Yang Z, Jang JC et al (2001) Modification of plant architecture in chrysanthemum by ectopic expression of the tobacco phytochrome B1 gene. J Am Soc Hort Sci 126:19–26

Zuker A, Tzfira T, Ben-meir H et al (2002) Modification of flower colour and fragrance by antisense suppression of the *flavanone 3-hydroxylase* gene. Mol Breed 9:33–41

Chapter 31
Transgenesis and Genomics in Forage Crops

Toshihiko Yamada, Ken-ichi Tamura, Xun Wang, and Yukiko Aoyagi

Abstract Forage grasses and legumes are important for animal production as well as soil conservation and environmental protection. Grass species also provide opportunities for turf and amenity. Further recently some of grasses showing high biomass have been given attention to renewable biomass energy crops. Many forage crops have polyploidy property and exhibit obligate outbreeding reproductive habits. Conventional breeding for forage crops is mainly based on synthetic varieties through polycrosses among superior selected genotypes. This system presents important challenges for the implementation of molecular breeding techniques such as transgenesis and genomics. Transgenesis and genomics approaches in forage crops have not been well performed due to many species and each species with relatively less importance compared to major cereals. However, recently transgenesis and genomics researches have been developed mainly for *Lolium* and *Festuca* species in grasses and *Triflolium* and *Medicago* species in legumes. In this charter current status of molecular breeding in forage crops is reviewed. Functionally-associated genetic markers which may be validated through association analysis will become important a tool for molecular breeding in near future. Author's group has research activities on genetic improvement of cold temperature tolerance in forage grasses. Some of our results on candidate gene approach are also described.

31.1 Introduction

Forage crops including grasses and legumes species have played an important role in providing pasture and hay for the livestock industry and are important in soil conservation and environmental protection. Besides being used as forage, some of the grass species, such as perennial ryegrass (*Lolium perenne* L.), tall fescue (*Festuca arundinacea* Schreb.), bent grass (*Agrostis* spp.), Kentucky bluegrass

T. Yamada (✉), X. Wang, and Y. Aoyagi
Field Science Center for Northern Biosphere, Hokkaido University, Sapporo, Japan
e-mail: yamada@fsc.hokudai.ac.jp

Ken-ichi Tamura
National Agricultural Research Center for Hokkaido Region, Sapporo, Japan

S.M. Jain and D.S. Brar (eds.), *Molecular Techniques in Crop Improvement*,
DOI 10.1007/978-90-481-2967-6_31, © Springer Science+Business Media B.V. 2010

(*Poa pratensis* L.) have been widely used for amenity purpose on lawns, parks, sports field, golf course and roadsides. In addition to food and environment, recently some of grasses showing high biomass such as switchgrass (*Panicum virgatum* L.) and reed canary grass (*Phalaris arundinacea* L.) have been given attention to renewable biomass energy crops as a possible alternative to fossil fuels. Otherwise forage legumes such as white clover (*Trifolium repens* L.), red clover (*Trifolium pratense* L.) and alfalfa (*Medicago sativa* L.) are important species, providing high quality herbage with nitrogen fixation. Forage legumes are essential for sustainable grasslands agriculture.

New forage crop cultivars which developed through genetic improvement have a long history of positively impacting forage and livestock systems. Traditional breeding methods such as phenotypic evaluation, selection and hybridization have always been, and still continue to be, used. However, forage crop breeding programs have entered the biotechnology era using molecular biology tools. Molecular breeding is therefore a relatively new term that describes the use of transgenic and genomic biotechnologies in a conjunction with traditional breeding programs (Bouton 2008).

Transgenesis involves the movement of specific and useful genes into the crop of choice. This approach has already been shown success in introducing genes which make many important crops resistant to insects, viruses, and herbicides. The transgenic approach has also been very useful in creating unique plants that allow basic research to clarify the physiological and biochemical pathways.

Genomics research received great publicity with the successful completion of the human genome sequencing project. After whole genome sequence of the model plant species, *Arabidopsis thaliana* (L.) Heynh. and *Oryza sativa* L., sequencing of several other plant species is currently underway; these include *Sorghum bicolor* Moench, *Brachypodium distachyon* (L.) P. Beauv., *Medicago truncatula* Gaertn. and *Lotus corniculatus* L. var. *japonicus* Regel, now being sequenced and used as a reference species for grasses and legumes, respectively. The sequencing data for these reference species, combined with high throughput machinery and data analysis (e.g. bioinformatics), allows more accurate determinations of species relationships and gene expression. From this understanding, new and innovative methods for improving forage crops are evolving.

31.2 Transgenesis

Genetic improvement of forage crops by conventional plant breeding is slow since many of forage plant species are predominantly, if not completely, allogamous wind-pollinated grasses and insect-pollinated legumes and they are generally self-sterile. Self-incompatibility limits inbreeding to concentrate desired genes for use in rapid development of new cultivars. Conventional forage crops breeding has been based on the use of natural genetic variation as found between and within ecotypes and cultivars or created through sexual recombination. Gene technology

and the production of transgenic plants offer the opportunity to generate unique genetic variation. Application of transgenesis to forage plant improvement has been focused on the development of transformation events with unique genetic variation and in studies on the molecular dissection of plant biosynthetic pathways and developmental processes of high relevance for forage production (Spangenberg et al. 2001).

31.2.1 Biolistic Transformation

Biolistic methodology, based on particle bombardment employs high-velocity gold or tungsten particles to deliver DNA into living cells for stable transformation (Sanford 1988; Christou 1992). Because biolistic methodology is a physical process that involves only one biological system, it is a fairy reproducible method that can be easily adapted from one laboratory to another laboratory. Transgenic forage plants have been obtained by particle bombardment of embryogenic cell in tall fescue (Spangenberg et al. 1995a; Cho et al. 2000; Wang et al. 2001, 2003a; Chen et al. 2003, 2004), perennial ryegrass (Spangenberg et al. 1995b; Dalton et al. 1999; Altpeter et al. 2000; Xu et al. 2001; Petrovska et al. 2004; Chen et al. 2005), Italian ryegrass (*Lolium mulitiflorum* Lam.) (Ye et al. 1997; Dalton et al. 1999; Ye et al. 2001; Takahashi et al. 2002, 2005; Li et al. 2004; Petrovska et al. 2004), orchardgrass (*Dactylis glomerata* L.) (Denchev et al. 1997; Cho et al. 2001), Kentucky bluegrass (Ha et al. 2001; Gao et al. 2006) as well as some warm season forage grasses such as switchgrass (Richards et al. 2001), bahiagrass (*Paspalum notatum* Flugge) (Gondo et al. 2005; Zhang et al. 2007a) and Rhodes grass (*Chloris gayana* Kunth) (Gondo et al. 2009).

31.2.2 Agrobacterium-Mediated Transformation

Agrobacterium-mediated transformation has the advantage of allowing for low copy number integration of the transgenes into the plant genome. In recent years, significant progress has been in developing transformation protocols using *Agrobacterium tumefaciens* as a vector. Transgenics have been obtained by *Agrobacterium*-mediated transformation in tall fescue (Dong and Qu 2005; Wang and Ge 2005a), perennial ryegrass (Altpeter et al. 2004; Sato and Takamizo 2006; Wu et al. 2005; Bajaj et al. 2006; Wu et al. 2007), Festulolium (*Lolium*/*Festuca* hybrids) (Guo et al. 2009), Italian ryegrass (Bettany et al. 2003) and orchardgrass (Lee et al. 2006) as well as warm season grasses such as switchgrass (Somleva et al. 2002, 2008) and zoysiagrass (*Zoysia japonica* Steud.) (Ge et al. 2006; Zhang et al. 2007b).

Compared with forage grasses, transformation is relatively easier in forage legume species. Efficient protocols for genetic transformation of the major forgae legumes using *Agrobacterium tumefaciens* have been established over last decade

(Kalla et al. 2001). For resent researches, transgenic alfalfa, white clover and *Medicago truncatula* plants have been obtained by *Agrobacterium*-mediated transformation (Barone et al. 2008; Rosellini et al. 2007; Montague et al. 2007; Balance et al. 2006; Crane et al. 2006; Wright et al. 2006; Xie et al. 2006; Zhang et al. 2005).

31.2.3 Selection Schemes

The establishment of efficient selection schemes by applying suitable selection pressure for an appropriate length of time is one of the critical aspects of successfully making transgenic plants. Because many forage grasses have high endogenous tolerance to antibiotics, particularly kanamycin, most of forage grasses transformation have been used *hph* or *bar* as selectable marker gene as and hygromycin or phosphinothricin (PPT) as selection agent (Wang and Yamada 2008).

Many reports of the transformation in forage grasses have been on method development using selectable marker genes and reporter genes (Wang and Yamada 2008). The selectable marker genes used are: hygromycin phosphotransferase gene (*hph*) from *E. coli*, phosphinothricin acetyltransferase gene (*bar*) from *Streptomyces hygroscopicus* and neomycin phosphotransferase II gene (*npt2*) of transposon Tn5. The reporter genes are β-glucuronidase gene (*gusA*) from *E. coli* and green fluorescent protein gene (*mgfp*) from *Aequorea victoria*. For *Agrobacterium*-mediated transformation, the *gusA* reporter gene needs to have an intron (e.g. a catalase intron) inside the coding sequence to ensure that expression of glucuronidase activity is derived from eukaryotic cells, not from expression by residual *A. tumefaciens* cells (Wang and Ge 2005a, b). The promoters used to construct the chimeric genes are: cauliflower mosaic virus (CaMV) 35S promoter, rice actin promoter, maize ubiquitin promoter or tissue specific promoters. The terminators used are the CaMV 35S terminator or the nos terminator from *Agrobacterium*.

31.2.4 Manipulation of Lignin Biosynthesis

Forage digestibility is a limiting factor for animal productivity. Lignification of plant well walls has been identified as the major factor responsible for lowering digestibility of forage tissues. Molecular breeding for improved digestibility by down-regulating monolignol biosynthetic enzymes through transgenesis has been explored. Forage digestibility of tall fescue has been improved by transgenic down-regulation of cinnamyl alcohol dehydrogenase (Chen et al. 2003) and caffeic acid O-methyltransferase (Chen et al. 2004).

Lignin also negatively affects the utilization of plant structural polysaccharides for ethanol production. Reducing lignin content may reduce recalcitrance to saccharification in cellulosic bioenergy crops such as switchgrass. Down-regulation of lignin in swithchgrass is being carried out using RNAi gene constructs (Wang et al. 2008).

31.2.5 Manipulation of Fructan Biosynthesis

Fructans, a polymer of fructose and a major component of nonstructural carbohydrates is accumulated in temperate grasses. The increased level of soluble carbohydrates appears to improve the nutritional value of grasses, particularly during summer when grasses suffer a major decline in digestibility. The soluble carbohydrate composition of Italian ryegrass has been altered by transformation with the *Bacillus subtilis sac B* gene (Ye et al. 2001). White clover does not normally accumulate fructan. Jenkins et al. (2002) have transgenic white clover plants that accumulate fructan, by expressing the fructosyltransferase (Ftf) enzyme from *Streptococcus salivarius* under the control of CaMV35S promoter. Instead of fructan synthesis genes from bacteria, Gadegaard et al. (2008) developed perennial ryegrass lines expressing sucrose:sucrose 1-fructosyltransferase and fructan:fructan 6G-fructosyltransferase genes from onion (*Allium cepa* L.) which exhibited up to a threefold increased fructan content. Fructan accumulation is also associated with winter hardiness. Transgenic perennial ryegrass plants that over-expressed the wheat fructosyltransferase genes, *wft1* and *wft2*, which encode sucrose-fructan 6-fructosyltransferase (6-SFT) and sucrose-sucrose 1-fructosyltransferase (1-SST), respectively, under the control of CaMV 35S promoter have been produced using a biolistic transformation (Hisano et al. 2004). Transgenic plants that accumulated a greater amount of fructan than non-transgenic plants showed increased tolerance to cellular freezing. The results suggest that the over-expression of the genes involved in fructan synthesis serves as a novel strategy to produce freezing-tolerant grasses (Hisano et al. 2004).

31.2.6 Gene Flow and Biosafety

Forage grasses are wind-pollinated and outcrossing species. Also forage legumes are insect-pollinated. Significant spread of the transgenes to other species or wild populations of the same species would be inevitable without strict control of pollination. Wang et al. (2004) reported pollen-mediated transgene flow using T_1 and T_2 progenies derived of transgenic tall fescue plants (Wang et al. 2003a, b). No transgene was detected at 200 m distance in any direction. This experiment indicated that the isolation distance 300 m is enough to prevent transgene flow to neighboring plants, at least small-scale field trials.

Because human indirectly consumes forage crops, biosafety evaluation of transgenic forage crops will likely to focus on their environmental or ecological impacts (Wang and Yamada 2008). Two questions need to be answered regarding biosafety of transgenic forage crops, first, how far can grass pollen disperse and still remain viable, and second, what is the probability of transgene escape by crossing with related species under natural conditions? Molecular markers technique could be powerful to answer two questions.

31.3 Genomics

Molecular breeding as molecular marker-assist selection could be powerful for the selection. In order to establish the molecular breeding, several researches on genome resources, DNA marker, genetic linkage map and quantitative trait loci (QTL) analysis and association analysis have been carried out in forage crops. Genomic studies in forage crops have not well been performed due to many species and each species with relatively less importance compared to major cereals. However, genomic researches have been developed mainly in the diploid species, perennial ryegrass, red clover and so on. Herbage yield and forage quality are important breeding objectives in a combination of several traits such as drought, heat and cold tolerance as well as disease and insect resistance. DNA marker selection approaches are expected to accelerate the conventional breeding approaches because agriculturally important traits showing continuous phenotypic variation are controlled by a variable number of QTL.

31.3.1 Genome Resources

31.3.1.1 ESTs

Gene discovery by expressed sequence tag (EST) sequencing has generated substantial genomic resources. In Australian group, a collection of 44,534 perennial ryegrass ESTs was generated from single pass sequencing of randomly selected clones from 29 cDNA libraries that represent a range of plant organs, developmental stages and environmental conditions (Sawbridge et al. 2003a). Similarly a collection of 42,017 white clover ESTs was generated from 16 cDNA libraries obtained from a broad range of plant organs, developmental stages and environmental conditions (Sawbridge et al. 2003b). Each of the sequences was annotated by comparison to GenBank and SwissProt public sequence databases and automated intermediate Gene Ontology (GO) annotation was obtained (Spangenberg et al. 2005). All sequences and annotation are maintained within ASTRA format MySQL databases, with web-based access for text searching, BLAST sequence comparison and GO hierarchical tree browsing. Each ryegrass sequence was mapped onto an EnsEMBL genome viewer for comparison with the complete genome sequence of rice and expressed sequences from related species. Similarly, each white clover sequence was mapped onto an EnsEMBL genome viewer for comparison with the complete *A. thaliana* genome sequence and expressed sequences from related legumes.

In *L. multiflorum*, 5,922 ESTs have been generated from seven cDNA libraries from various tissues and leaves under biotic and abiotic stresses (Ikeda et al. 2004). A gene inventory of 7,810 unique gene clusters has been generated from a total of 11,990 individual sequences using cDNA libraries of switchgrass (Tobias et al. 2005). A total of 7,029 unigenes have been functionally categorized from EST collection of semiarid native grass, *Eragrostis curvula* (Schrad.) Nees (Cervigni et al. 2008). A total of 26,356 ESTs of red clover were collected, and 78% of the ESTs showed sequence similarity to registered genes, mainly of *Arabidopsis thaliana* and rice (Sato et al. 2005).

31.3.1.2 BAC Libraries

Complementing the EST resources, large insert DNA libraries have been generated using bacterial artificial chromosome (BAC) vectors. Two BAC libraries were constructed for perennial ryegrass (Farrar et al. 2007). The libraries consisted of 98,304 and 101,376 BAC clones for perennial ryegrass genotypes LTS18 and NV#20F1-30, respectively. The estimated average insert size of both libraries was approximately 100 kb. BAC libraries of perennial ryegrass (50,304 BAC clones with 113 kb average insert size, corresponding to 3.4 genome equivalents and 97% genome coverage) and white clover (50,302 BAC clones with 101 kb average insert size, corresponding to 6.3 genome equivalents and 99% genome coverage) have been established (Spangenberg et al. 2005).

31.3.2 Transcriptomics

High-density cDNA microarrays representing approximately 15,000 unique genes for each of perennial ryegrass and white clover (Sawbridge et al. 2003a, b) have been developed. The forage crops-derived microarrays have been applied in hybridizations with labelled total RNA isolated from a variety of genotypes, plant organs, developmental stages, and growth conditions. Results from these studies have enabled validation of functions predicted through comparative sequence annotation, and also suggested roles for novel genes which lack comparative sequence annotation (Spangenberg et al. 2005). Unannotated genes which are co-regulated with the perennial ryegrass *LpCAD1* and *LpCAD3* (cinnamyl alcohol dehydrogenase) lignin biosynthesis genes have been identified and mapped as candidate gene-based markers to regions of the genome associated with herbage QTL (Cogan et al. 2006b). Microarrays may also be used for gene and promoter discovery when used in concert with the BAC libraries established for each species (Spangenberg et al. 2005). An International Transcriptome Initiative for Forage and Turf (ITIFT) has been proposed to facilitate international efforts in microarray-based transcriptome analyses for important forage crops (Spangenberg et al. 2005).

31.3.3 DNA Markers

31.3.3.1 SSR Markers

Simple sequence repeats (SSR) markers provide the current marker system of choice due to their abundance, ubiquitous distribution in plant genomes, high level of reproducibility, ease of PCR-based analysis, and detection of co-dominant multiallelic loci. Thus, SSR markers have been developed in forage crops such as perennial ryegrass (Jones et al. 2001; Kubik et al. 2001; Asp et al. 2007; Jensen et al. 2007; Studer et al. 2008), Italian ryegrass (Hirata et al. 2006; Cai et al. 2008),

tall fescue (Saha et al. 2004, 2006), timothy (*Phleum pratense* L.) (Cai et al. 2003), zoysiagrass (Cai et al. 2005a, b), orchardgrass (Cai et al. 2008) and switchgrass (Tobias et al. 2006). SSR markers for forage legumes have been also reported in white clover (Kölliker 2001; Jones et al. 2003; Barrett et al. 2004), red clover (Sato et al. 2005) and alfalfa (Diwan et al. 2000; Sledge et al. 2005).

31.3.3.2 DNA Markers from Rice Genome Information

Marker systems developed using rice genomic information (International Rice Genome Project 2005) were proposed for related Poaceae crops (Lem and Lallemand 2003; Fredslund et al. 2006; Ishikawa et al. 2007). These systems developed PCR-based markers from cDNA sequences, including ESTs, of the related species showing high similarity to single-copy rice genes (with no paralogous genes). Thus, comparative genomic analyses allow the genomic loci of markers in the related species to be estimated from the corresponding the rice loci. In this marker system, primer design is based on the sequences of exonic regions flanking introns of rice genes (Fig. 31.1). Because intronic regions generally include more polymorphisms than exonic regions, a high frequency of polymorphic markers is expected using this system. Novel PCR-based EST markers were developed all designed around intronic regions which show higher polymorphism than exonic regions. Intronic regions of the grass genes were speculated using rice genomic information. Two hundred and nine primer sets were designed from *Lolium/Festuca* ESTs showing high similarity to unique rice genes dispersed uniformly throughout the rice genome. Sixty one of these primer sets as insertion-deletion

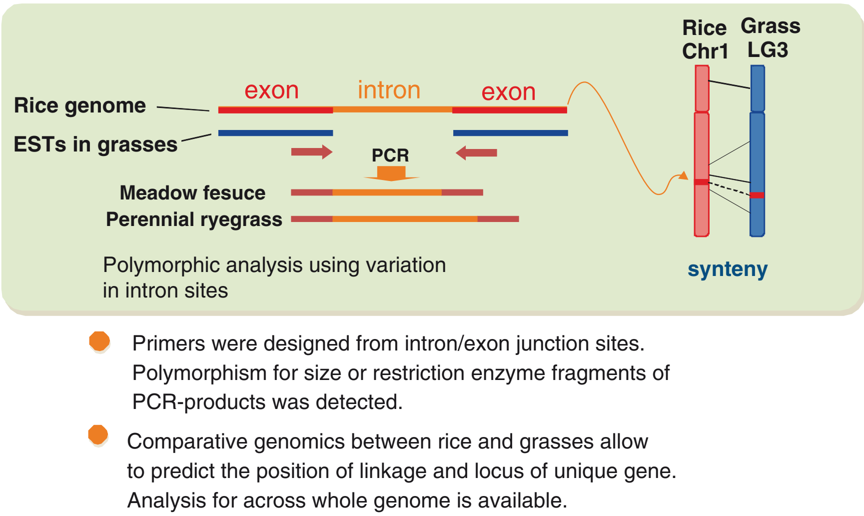


Fig. 31.1 Intron-flanking EST markers based on comparative genomics between *Lolium perenne*/*Festuca pratensis* and rice genome

(indel) -type markers and 82 primer sets as cleaved amplified polymorphic sequence (CAPS) markers to distinguish between perennial ryegrass and meadow fescue (*Festuca pratensis* Huds) (Tamura et al. 2009). Many indel-type markers and CAPS markers had high species specificity to *L. perenne* and *F. pratensis*. Chromosome mapping of these markers using *Lolium/Festuca* substitution plants revealed syntenic relationships between *Lolium/Festuca* and rice largely consistent with previous reports (Tamura et al. 2009). This marker system based on intron polymorphisms with a high frequency between species and high species specificity could consequently be a useful tool for breeding of Festulolium, intergeneric hybrid between *Lolium* and *Festuca*.

31.3.4 Genetic Map

The development of molecular genetic markers and associated genetic maps for forage crops has been comprehensively reviewed by Forster et al. (2004). Many forage crops have not well been established precise genetic linkage maps because of outcrossing property. However, linkage maps have been constructed using SSR markers in perennial ryegrass (Jones et al. 2002b; Gill et al. 2006), Italian ryegrass (Hirata et al. 2006), tall fescue (Saha et al. 2005), red clover (Sato et al. 2005), white clover (Barrett et al. 2004; Zhang et al. 2007c) and alfalfa (Sledge et al. 2005).

An enhanced molecular marker-based genetic linkage map of perennial ryegrass has been constructed through the activities of the International *Lolium* Genome Initiative (ILGI) (Forster et al. 2001), using the p150/112 one-way pseudo-testcross mapping population. The map contains 109 RFLP loci detected by heterologous probes from wheat, barley, oat and rice. Comparative genetic mapping has allowed the alignment of the perennial ryegrass genetic map with those of wheat, rice and oat, revealing substantial conserved synteny with the genomes of Poaceae species (Jones et al. 2002a). At the macrosyntenic level, each of the seven linkage groups (LGs) of perennial ryegrass chiefly corresponds to one of the seven basic homeologous chromosome groups of the Triticeae cereals, and they have been numbered accordingly. Seven LGs of perennial ryegrass also correspond to 12 LGs of rice.

Two genetic mapping populations of perennial ryegrass (F_2 (Aurora x Perma) and F_1 (North African$_6$ (NA_6) x Aurora$_6$ (AU_6)) have been independently developed as successors to the p150/112 population, and have been aligned to the reference map using common markers (Armstead et al. 2004; Faville et al. 2004). These genetic maps contain functionally-associated molecular marker information through the inclusion of gene-associated CAPS markers, and both RFLP and SSR markers from ESTs, respectively. In addition, high- density molecular marker-based genetic linkage maps have also been constructed for Italian ryegrass (Inoue et al. 2004a; Hirata et al. 2006), meadow fescue (Alm et al. 2003), tall fescue (Saha et al. 2005), interspecific hybrid populations based on crosses between *L. multiflorum* and *L. perenne* (Warnke et al. 2004; Sim et al. 2005), zoysiagrass (Cai et al. 2004,

2005), timothy (Cai et al. 2008) and orchardgrass (Cai et al. 2008). Genetic maps are consequently available for detailed dissection of complex phenotypes to resolve the locations of pleiotropic and interacting genetic factors.

A comprehensive diploid red clover genetic map has been constructed using the F_1 two-way pseudo-testcross population, initially through the use of 157 cDNA-derived RFLP markers (Isobe et al. 2003), and more recently through incorporation of genomic DNA-derived and EST-SSR markers to generate a map containing 1,434 SSR loci across the 7 LGs with a cumulative map length of 869 cM (Sato et al. 2005). Macrosynteny was determined between red clover and both *Medicago truncatula* and *Lotus japonicus*. A second two-way pseudo-testcross population has been used to construct a map containing 216 AFLP and 42 SSR loci, with a cumulative length of 444 cM (Herrmann et al. 2006).

Genetic map development in allotetraploid white clover was performed using a combination of SSR and AFLP markers. The F_2 genetic map contained 135 loci (78 SSR and 57 AFLP) on 18 LGs (two more than the karyotypic number), with a total map length of 825 cM (Jones et al. 2003). A higher-resolution genetic map largely based on EST-SSR markers was constructed using the F_1 mapping family (Barrett et al. 2004). A total of 335 EST-SSR and 30 genomic DNA-derived SSR primer pairs detected polymorphism and permitted assignment of 493 loci to a genetic map containing 16 LGs, with a total map length of 1,144 cM. The EST-SSR markers detected homoeologous locations between the ancestral genomes at high frequency, and provided the basis for standard chromosome nomenclature development. Recently third genetic linkage map of white clover was constructed using SSR markers from several Trifolieae species, including white clover, red clover, *Medicago truncatula* and soybean using F_1 mapping population (Zhang et al. 2007c). LGs for all eight homoeologous chromosome pairs of white clover were detected. Map length was estimated at 1,877 cM with 87% genome coverage.

Genomic DNA-derived and EST-SSR markers have been used to construct a functionally-associated marker-based tetraploid alfalfa genetic map (Sledge et al. 2005) in a pair-cross population varying for drought tolerance-associated traits.

31.3.5 Trait Dissection

The application of genetic marker analysis to trait dissection has been reviewed by Yamada and Forster (2005) and Yamada et al. (2005). Trait dissection for perennial ryegrass has been performed in multiple populations to allow QTL analysis. The p150/112 population has been analyzed for traits such as vegetative and reproductive morphogenesis, reproductive development and winter-hardiness, herbage quality, and gametophytic self-incompatibility (Thorogood et al. 2002; Yamada et al. 2004; Cogan et al. 2005; Shinozuka et al. 2006), while the $F_1(NA_6 \times AU_6)$ population has been studied for a range of root and shoot morphogenesis, photosynthetic efficiency, pseudostem water soluble carbohydrate content, and crown rust resistance

characters (Forster et al. 2004). Other perennial ryegrass populations have been analyzed to detect genetic control of crown rust resistance (Dumsday et al. 2003; Muylle et al. 2005a, b), vernalization response (Jensen et al. 2005) and flowering time variation (Armstead et al. 2004). Interspecific hybrid populations between *L. multiflorum* and *L. perenne* have been used to identify QTLs for flowering time variation (Warnke et al. 2004), fiber components and crude protein content (Xiong et al. 2007a), winter hardiness (Xiong et al. 2007b), gray leaf spot resistance (Curley et al. 2005) and crown rust resistance (Sim et al. 2007).

Trait-dissection in Italian ryegrass has permitted the detection of QTLs for lodging resistance and related traits such as heading date, plant height, culm weight, culm diameter, culm strength and tiller number (Inoue et al. 2004b). EST-derived CAPS and AFLP markers were also used to map and tag a major gene locus (*LmPil*) for ryegrass blast (grey leaf spot) resistance (Miura et al. 2005). A large two-way pseudo-testcross F_1 progeny set based on two highly heterozygous individuals with contrasting resistance to bacterial wilt (*Xanthomonas translucens* pv. *graminis*) has been used to generate an SSR- and AFLP-based map and to locate a single large (accounting for 67% of total phenotypic variance) resistance QTL (Studer et al. 2006). Two major QTLs for crown resistance were consistently detected on LG 1 and LG 2 in Italian ryegrass, explaining up to 56% of total phenotypic variance. Nevertheless, differences between position and magnitude of QTLs were observed among individual field locations and suggested the existence of specific local pathogen populations (Studer et al. 2007). The meadow fescue mapping population has been used for QTL analysis of vernalization requirement, heading time and number of panicles (Ergon et al. 2006), and seed production traits (Fang et al. 2004).

The F_1 genetic map of red clover containing 216 AFLP and 42 SSR loci was exploited for QTL analysis of seed yield component traits, identifying 38 QTLs for 8 traits, including 2 genomic regions with clustered QTLs suitable for marker-based selection (Herrmann et al. 2006). Persistency which is important trait for red clover has been also identified using same mapping population (Herrmann et al. 2008).

The F_2 genetic map of white clover has been exploited for QTL analysis of a number of vegetative morphogenesis, reproductive morphogenesis and reproductive development traits (Cogan et al. 2006a). Target traits were measured across a number of years of clonal replication, and geographical sites in Wales and Scotland, United Kingdom. Individual environment analyses detected a large number of QTLs for each trait, with QTL clustering for correlated traits, especially on LGs 7 and 12. Multi-environment combined analysis revealed genomic locations that are relatively insensitive to genotype x environment effects. The F_1 population has also been used for QTL analysis, specifically targeting seed production traits such as inflorescence density, yield per inflorescence and thousand-seed weight (Barrett et al. 2005). Stability of QTL effects was observed across temporal replication, along with co-location of QTLs for correlated traits.

QTLs controlling forage production, forage height, and forage regrowth at three harvests in tetraploid alfalfa were mapped on a genetic linkage map developed from hybrids between *Medicago sativa* subsp. *falcata* and *M. sativa* subsp. *sativa*. (Robins

et al. 2007b). The most marker-trait associations were detected during the first harvest. Markers associated with autumn regrowth tended to be different from those associated with regrowth earlier in the season, suggestive of the identification of autumn dormancy effects. Robins et al. (2007a) evaluated tetraploid alfalfa population for biomass production over several years at three locations. Despite the complexities of mapping within autotetraploid populations, single-marker analysis of variance identified 41 marker alleles, many on LGs 5 and 7, associated with biomass production in at least one of the sampling periods. Seven alleles were associated with biomass production in more than one of the sampling periods. QTL involved in resistance and susceptibility to *Stagonospora meliloti* (Musial et al. 2007) and reaction to three races of *Colletotrichum trifolii* (Mackie et al. 2007) were identified in an autotetraploid alfalfa backcross population of 145 individuals. Other trait dissection studies in autotetraploid alfalfa have identified resistance to root rot (*Phytophthora medicaginsis*) (Musial et al. 2005) and yield, yield components and morphological traits (Musial et al. 2006). Three putative aluminum tolerance QTLs on LG I, LG II and LG III, explaining 38%, 16% and 27% of the phenotypic variation, respectively, were identified in diploid alfalfa (Narasimhamoorthy et al. 2007). Six candidate gene markers designed from *Medicago truncatula* ESTs that showed homology to known Al-tolerance genes were placed on the QTL map.

The phenotypic variation of a complex trait usually results from multiple QTLs, QTL–QTL interactions, and QTL–environmental interactions. In order to reveal QTL interactions and the relationship between various interactions in complex traits, a new QTL mapping approach, named genotype matrix mapping (GMM), which searches for QTL interactions in genetic variation has been developed (Isobe et al. 2007). GMM will be a valuable approach to identify QTL interactions in genetic variation of a complex trait within a variety.

31.3.6 Marker-Assist Selection

Genetic gain from phenotypic selection in open-pollinated forage crops is constrained by the inability to accurately use phenotype to estimate genotype, prior to parent selection for polycrossing. The use of marker-assisted selection (MAS) offers the potential to accelerate genetic gain by overcoming this constraint. SSR marker associated with QTLs for seed production allowed to be realized the value of MAS in white clover (Barrent et al. 2005, 2008). Identifying markers based on bi-parental mapping populations is likely not the best way to implement a MAS program, although this approach is useful to introgress alleles from wild germplasm (Brummer and Casler 2008). Instead, a more practical approach may be the use of association analysis, measuring both phenotypes and markers directly on the plants in the breeding nursery.

31.3.7 Association Analysis

Marker-trait gene linkage analysis method is not readily used for outbreeding forage crops. Functionally-associated candidate genes have been developed in some forage crops. Forage crops would have a potentiality to dispose towards limited linkage disequilibrium (LD), extending over relatively short molecular distances (Forster et al. 2004; Dobrowolski and Forster 2007). Forage crops contain many accessions with long-established populations derived from a large number of founding parents, as expected for ecotypes and long-established varieties, in which many rounds of recombination have occurred. These accessions allow to use candidate gene-based functionally-associated marker analyses (Andersen and Lübberstedt 2003). Nucleotide variation in qualified candidate genes will be closely associated with the casual mutations that generate variability for key agronomic traits, and may be used as diagnostics for such variation. Successful correlation of gene haplotype structure and phenotypic variation will provide the basis for a new paradigm in forage crop molecular breeding based on direct selection of superior allele content at target genetic loci, allowing highly effective exploitation of germplasm collections for identification of potential parental genotypes (Forster et al. 2004; Spangenberg et al. 2005). LD was investigated in the gibberelic acid insensitive gene region in three synthetic varieties of perennial ryegrass chosen for their contrasted number of parents in the initial polycrosses using Sequence Tagged Site (STS) and SSR markers (Auzanneau et al. 2007). Significant LDs were observed up to 1.6 Mb in a variety originated from 6 related parents and not above 174 kb in a variety originated from 336 parents, suggesting that an association analysis can be used when varieties have a large number of parents.

Single nucleotide polymorphisms (SNPs) are the most versatile class of functionally associated genetic marker. Efficient methods for *in vitro* discovery of SNPs and characterization of SNP haplotype structure have been described for perennial ryegrass (Spangenberg et al. 2005; Cogan et al. 2006b; Ponting et al. 2007). SNP discovery in perennial ryegrass has been based on PCR amplification and sequencing of multiple amplicons designed to scan all components of the transcriptional unit. Multiple SNPs at regular intervals across the transcriptional unit were detected within and between the heterozygous parents and validated in the progeny of the $F_1(NA_6 \text{ x } AU_6)$ genetic mapping family. Haplotype structures in the parental genotypes were defined and haplotypic abundance, structure and variation were assessed in diverse germplasm sources. Decay of LD to r^2 values of c. 0.2 typically occurs over 500–3,000 bp, comparable with gene length and with little apparent variation between diverse, ecotypic and varietal population sub-groups (Smith et al. 2008).

A candidate gene approach for associating SNPs with variation in flowering time and water-soluble carbohydrate content (WSC) and other traits has been described for perennial ryegrass (Skøt et al. 2007). Analysis of variance within populations identified several associations between WSC, nitrogen, and dry matter

digestibility with allelic variants within an alkaline invertase candidate gene *LpcAI*. Consistent associations between the perennial ryegrass homolog (*LpHD1*) of the rice photoperiod control gene *HD1* and flowering time were identified. One SNP, in the immediate upstream region of the *LpHD1* coding sequence (C-4443-A), was significant in the linear mixed model.

An EST resource obtained from multiple cDNA libraries constructed from numerous genotypes of a single cultivar has been used for *in silico* SNP discovery and validation in white clover. A total of 58 from 236 selected sequence clusters (24.5%) were fully validated as containing polymorphic SNPs by genotypic analysis across the parents and progeny of several two-way pseudo-testcross mapping families of white clover (Cogan et al. 2007).

31.4 Candidate Gene Approach – Cold Responsible Genes

Author's group has researched activities on genetic improvement of cold temperature tolerance in perennial ryegrass. Many plants exhibit an increase in freezing tolerance in response to low, non-freezing temperatures, a phenomenon known as cold acclimation. A number of genes respond to cold and condition the plant cells against the effect of freezing temperature during cold acclimation. Some of our results on candidate gene approach have been described as follows.

31.4.1 CBF Genes

Molecular mechanisms involved in cold acclimation are largely unknown but information from model species whose genomes have recently been sequenced such as Arabidopsis and rice and the development of microarray technologies are giving insight into the complexity of the processes (Yamaguchi-Shinozaki and Shinozaki 2006, 2008). The CBF (C-repeat binding factor) /DREB1 (dehydration-responsive element-binding protein 1) regulon is the most important transcription unit involved in cold acclimation in plants (Nakashima and Yamaguchi-Shinozaki 2006). Ten novel putative *CBF* cDNAs have been isolated from cold-treated leaf tissue of perennial ryegrass (Tamura and Yamada 2007). Their primary structures contain some conserved motifs characteristic of the gene class. Phylogenetic analysis revealed that *LpCBF* genes were attributable to the HvCBF3-, and HvCBF4-subgroups following the previously proposed classification of barley *CBF* genes (Skinner et al. 2005). RT-PCR analysis revealed that the expression of *LpCBF* genes was rapidly induced in response to low temperature and that the expression pattern under the low temperature conditions for a long period was different between the various *LpCBF* genes. Five of the ten *LpCBF* genes were assigned to the genetic linkage map. Four *LpCBF* genes were mapped on LG 5 forming a cluster within 2.2 cM, while one *LpCBF* gene on LG 1. Based on comparative genetic

studies, conserved synteny for *CBF* gene family was observed between the Triticeae cereals and perennial ryegrass (Tamura and Yamada 2007). *CBF* genes cluster is positioned at the frost resistance locus, *Fr-H2* in barley (Francia et al. 2007). Freezing tolerance in tillers was also evaluated and QTL for this trait was observed on LGs 5 and 6 (Yamada et al., unpublished). Determination of the functional role of each different expression type *LpCBF* gene will be necessary for development of specific genetic markers associated with the low-temperature tolerance.

31.4.2 Fructosyltransferase Genes and Fructan QTL

Fructans are linear or branched forms of fructose polymers, which are derived from sucrose. Fructans are present in 15% of the angiosperm flora, and are particularly widespread in grasses (Chatterton et al. 1989; Hendry 1993). Fructans accumulate in plant cells as a carbohydrate reserve in addition to or instead of starch (Hendry 1993), and are also thought to be involved in the maintenance of osmotic potentials (Pavis et al. 2001). Accumulation of fructans in plants has been found to be associated with tolerance to cold and drought, particularly in the development of freezing tolerance (Yoshida et al. 1998; Kawakami and Yoshida 2002). Fructan is synthesized by a combination of multiple fructosyltransferases (FTs). Two *FT* genes have been identified in perennial ryegrass: *1-SST* (Chalmers et al. 2003) and *6G-FFT* (Lasseur et al. 2006). FT genes have been mapped to perennial ryegrass LGs using a mapping population derived from a pair-cross between North $African_6$ (NA_6) and $Aurora_6$ (AU_6) plants (Chalmers et al. 2005). The 1-SST gene (*Lp1-SST*; marker xlp1-sst) has been mapped to a distal region of LG7 of NA_6.

Six cDNAs encoding FTs (*prft1*-*prft6*) have been isolated from cold-treated perennial ryegrass plants (Hisano et al. 2008). The *prft1* and *prft4* genes were both located at near a gene for soluble invertase in the distal part of the LG 7 using F_2 (Aurora x Perma) mapping population in Armstead et al. (2004). The *prft3* gene was located in the distal part of the LG 3. Functional characterization using *Pichia pastoris* revealed that the *prft4* encodes sucrose-sucrose 1-fructosyltransferase (1-SST), and the *prft3* and *prft5* encode fructan-fructan 6G-fructosyltransferase (6G-FFT). Protein sequences for the other genes (*prfts 1*, *2*, and *6*) were similar to sucrose-fructan 6-fructosyltransferase (6-SFT). The mRNA levels of *prft1* and *prft2* gradually increased during cold treatment while those of the 1-SST and 6G-FFT genes first increased but then decreased before increasing again during a longer period of cold treatment. At least two different patterns of expression of FT genes appear to have developed during the evolution of functionally diverse FT genes which are associated in a coordinated way with fructan synthesis in a cold environment (Hisano et al. 2008).

The F_2 (Aurora x Perma) mapping population was used to identify QTL for traits relating to winter hardiness as well as sugar content (Yamada et al., unpublished). Scores of winter survival were measured in the deep snow environment at Sapporo using fungicide control regime of snow mould fungi, and QTL were identified on

LGs 2, 4, 6 and 7. QTL for content of high molecular fructan were observed on LGs 1, 2 and 4. QTL for winter survival in LGs 2 and 4 are close to QTL for fructan content. Freezing tolerance in tillers was also evaluated and QTL for this trait was observed on LGs 5 and 6. QTL for sucrose content on LG 5 was closed to one QTL for freezing tolerance. Fructosyltransferase genes were mapped to the genetic map, but failed to show coincidence with any fructan content QTL. QTLs for fructans and the other component of WSC (sucrose, glucose and fructose) in leaves and tiller bases have been also mapped using F_2 (Aurora x Perma) mapping population (Turner et al. 2006). Fructan QTL were identified on chromosome 1, 2, 5 and 6. Many QTLs for growth-trait and drought-stress response were also identified on all chromosomes except for chromosome 7 (Turner et al. 2008). Therefore there is currently little evidence for FT involvement in these fructan QTL. Further isolation of FT genes and mapping may show closer association. Alternatively the QTL could result from variation in furctan hydrolases or from the activity of regulatory genes (Turner et al. 2006).

31.4.3 Other Genes

The α-subunit of the casein protein kinase CK2 has been implicated in both light-regulated and circadian rhythm-controlled plant gene expression, including control of flowering time. Two putative CK2 α genes of perennial ryegrass (*Lpck2a-1* and *Lpck2a-2*) have been obtained from a cDNA library constructed with mRNA isolated from cold-acclimated crown tissue (Shinozuka et al. 2005). The *Lpck2a-1* CAPS marker was assigned to perennial ryegrass LG 4 and the *Lpck2a-2* CAPS marker was assigned to LG2. Allelic variation at the *Lpck2a-1* and *Lpck2a-2* gene loci was correlated with phenotypic variation for heading date and winter survival, respectively (Shinozuka et al. 2005). The gene for a putative glycine-rich RNA binding protein, *LpGRP1*, was isolated from a cDNA library constructed from crown tissues of cold-treated perennial ryegrass plants (Shinozuka et al. 2006). An RFLP locus detected by the *LpGRP1* cDNA probe was mapped to a distal location on LG 2 in the p150/112 population. A significant increase in the mRNA level of *LpGRP1* was detected in root, crown and leaf tissues during the treatment of plants at 4°C, through which freezing tolerance is attained. LpGRP1 protein could play an important role for adaptation to cold environments.

31.5 Conclusion

Molecular breeding is important and will be used extensively in future forage crop improvement and development of biofuel crops as cellulosic biomass feedstock. Transformation techniques using *Agrobacterium* or biolistics-based method and many available molecular markers such as SSR markers and some functionally-associated

genetic markers have been developed in many forage crops. We have still remaining challenge on the successful implementation of molecular breeding in practical varietal development. Molecular breeding needs to develop from a platform of good conventional breeding and include supporting agronomic research and partnering with commercial industry where appropriate. With the availability of more sequencing information such ESTs, gene isolation has much easier than ever before. We should focus on functional characterization of genes and their regulatory elements.

With a widening range of traits, techniques for more accurate, rapid and non-invasive phenotyping and genotyping become increasingly important. The large amounts of data involved require good bioinfomatics support. Data of various kinds must be integrated from an increasingly wide range of sources such as genetic resources and mapping information for plant populations through to the transcriptome and metabolome of individual tissues. The merging of data from disparate sources and multivariate data-mining across datasets can reveal novel information concerning the biology of complex systems.

References

Alm V, Fang C, Busso CS, Devos KM, Vollan K, Grieg Z, Rognli OA (2003) A linkage map of meadow fescue (*Festuca pratensis* Huds.) and comparative mapping with other Poaceae species. Theor Appl Genet 108: 25–40

Altpeter F, Xu JP, Ahmed S (2000) Generation of large numbers of independently transformed fertile perennial ryegrass (*Lolium perenne* L.) plants of forage- and turf-type cultivars. Mol Breed 6:519–528

Altpeter F, Fang YD, Xu JP, Ma XR (2004) Comparison of transgene expression stability after *Agrobacterium*-mediated or biolistic gene transfer into perennial ryegrass (*Lolium perenne* L.). In: Hopkins A, Wang ZY, Mian R, Sledge M, Barker RE (eds) Molecular breeding of forage and turf. Kluwer Academic, Dordrecht, pp 255–260

Andersen JR, Lübberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554–560

Armstead IP, Turner LB, Farrell M, Skøt L, Gomez P, Montoya T, Donnison IS, King IP, Humphreys MO (2004) Synteny between a major heading-date QTL in perennial ryegrass (*Lolium perenne* L.) and the Hd3 heading-date locus in rice. Theor Appl Genet 108:822–828

Asp T, Frei UK, Didion T, Nielsen KK, Lübberstedt T (2007) Frequency, type, and distribution of EST-SSRs from three genotypes of *Lolium perenne*, and their conservation across orthologous sequences of *Festuca arundinacea*, *Brachypodium distachyon*, and *Oryza sativa*. BMC Plant Biol 7:36

Auzanneau J, Huyghe C, Julier B, Barre P (2007) Linkage disequilibrium in synthetic varieties of perennial ryegrass. Theor Appl Genet 115:837–847

Bajaj S, Ran Y, Phillips J, Kularajathevan G, Pal S, Cohen D, Elborough K, Puthigae S (2006) A high throughput *Agrobacterium tumefaciens*-mediated transformation method for functional genomics of perennial ryegrass (*Lolium perenne* L.). Plant Cell Rep 25:651–659

Balance C-MC, McManus MT (2006) Expression of 1-aminocyclopropane-1-carboxylate (ACC) oxidase genes during the development of vegetative tissues in white clover (*Trifolium repens* L.) is regulated by ontological cues. Plant Mol Biol 60:451–467

Barone P, Rosellini D, LaFayette P, Bouton J, Veronesi F, Parrot W (2008) Bacterial citrate synthase expression and soil aluminum tolerance in transgenic alfalfa. Plant Cell Rep 27:893–901

Barrett B, Griffiths A, Schreiber M, Ellison N, Mercer C, Bouton J, Ong B, Forster J, Sawbridge T, Spangenberg G, Bryan G, Woodfield D (2004) A microsatellite map of white clover (*Trifolium repens* L.). Theor Appl Genet 109:596–608

Barrett BA, Baird IJ, Woodfield DR (2005) A QTL analysis of white clover seed production. Crop Sci 45:1844–1850

Barrett BA, Baird IJ, Woodfield DR (2008) White clover seed yield: a case study in marker-assisted selection. In: Yamada T, Spangenberg G (eds) Molecular breeding of forage crops. Springer Science + Business Media, New York, pp 241–250

Bettany AJE, Dalton SJ, Timms E, Manderyck B, Dhanoa MS, Morris P (2003) *Agrobacterium tumefaciens*-mediated transformation of *Festuca arundinacea* (Schreb.) and *Lolium multiflorum* (Lam.). Plant Cell Rep 21:437–444

Bouton JH (2008) Molecular breeding to improve forages for use in animal and biofuel production systems. In: Yamada T, Spangenberg G (eds) Molecular breeding of forage crops. Springer Science + Business Media, New York, pp 1–13

Brummer EC, Casler MD (2008) Improving selection in forage, turf, and biomass crops using molecular markers. In: Yamada T, Spangenberg G (eds) Molecular breeding of forage crops. Springer Science + Business Media, New York, pp 193–209

Cai H-W, Yuyama N, Tamaki H, Yoshizawa A (2003) Isolation and characterisation of simple sequence repeat markers in the hexaploid forage grass timothy (*Phleum pratense* L.). Theor Appl Genet 107:1337–1349

Cai H, Inoue M, Yuyama N, Nakayama S (2005a) An AFLP-based linkage map of Zoysiagrass (*Zoysia japonica*). Plant Breed 123:543–548

Cai H, Inoue M, Yuyama N, Takahashi W, Hirata M, Sasaki T (2005b) Isolation, characterization and mapping of simple sequence repeat markers in zoysiagrass (*Zoysia* spp.). Theor Appl Genet 112:158–166

Cai H-W, Inoue M, Yuyama N, Hirata M (2008) Genome mapping in cool-season forage grasses. In: Yamada T, Spangenberg G (eds) Molecular breeding of forage crops. Springer Science + Business Media, New York, pp 173–192

Cervigni GDL, Paniego N, M Di'az, Selva JP, Zappacosta D, Zanazzi, Landerreche I, Martelotto L, Felitti S, Pessino S, Spangenberg G, Echenique V (2008) Expressed sequence tag analysis and development of gene associated markers in a near-isogenic plant system of *Eragrostis curvula*. Plant Mol Biol 67:1–10

Chalmers J, Johnson X, Lidgett A, Spangenberg G (2003) Isolation and characterisation of a sucrose: sucrose 1-fructosyltransferase gene from perennial ryegrass (*Lolium perenne* L.). J Plant Physiol 160:1385–1391

Chalmers J, Lidgett A, Cummings N, Cao Y, Forster J, Spangenberg G (2005) Molecular genetics of fructan metabolism in perennial ryegrass. Plant Biotechnol J 3:459–474

Chatterton NJ, Harrison PA, Bennett JH, Asay KH (1989) Carbohydrate partitioning in 185 accessions of Gramineae under warm and cool temperatures. J Plant Physiol 134:169–179

Chen L, Auh C, Dowling P, Bell J, Chen F, Hopkins A, Dixon RA, Wang Z-Y (2003) Improved forage digestibility of tall fescue (*Festuca arundinacea*) by transgenic down-regulation of cinnamyl alcohol dehydrogenase. Plant Biotechnol J 1:437–449

Chen L, Auh C, Dowling P, Bell J, Lehmann D, Wang Z-Y (2004) Transgenic down-regulation of caffeic acid O-methyltransferase (COMT) led to improved digestibility in tall fescue (*Festuca arundinacea*). Funct Plant Biol 31:235–245

Chen X, Yang WQ, Sivamani E, Bruneau AH, Wang BH, Qu RD (2005) Selective elimination of perennial ryegrass by activation of a pro-herbicide through engineering *E. coli argE* gene. Mol Breed 15:339–347

Cho MJ, Ha CD, Lemaux PG (2000) Production of transgenic tall fescue and red fescue plants by particle bombardment of mature seed-derived highly regenerative tissues. Plant Cell Rep 19:1084–1089

Cho MJ, Choi HW, Lemaux PG (2001) Transformed T_0 orchardgrass (*Dactylis glomerata* L.) plants produced from highly regenerative tissues derived from mature seeds. Plant Cell Rep 20:318–324

Christou P (1992) Genetic transformation of crop plants using microprojectile bombardment. Plant J 2:275–281

Cogan NOI, Smith KF, Yamada T, Francki MG, Vecchies AC, Jones ES, Spangenberg GC, Forster JW (2005) QTL analysis and comparative genomics of herbage quality traits in perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 110:364–380

Cogan NOI, Abberton MT, Smith KF, Kearney G, Marshall AH, Williams A, Michaelson-Yeates TPT, Bowen C, Jones ES, Vecchies AC, Forster JW (2006a) Individual and multi-environment combined analyses identify QTLs for morphogenetic and reproductive development traits in white clover (*Trifolium repens* L.). Theor Appl Genet 112:1401–1415

Cogan NOI, Ponting RC, Vecchies AC, Drayton MC, George J, Dobrowolski MP, Sawbridge TI, Spangenberg GC, Smith KF, Forster JW (2006b) Gene-associated single nucleotide polymorphism (SNP) discovery in perennial ryegrass (*Lolium perenne* L.). Mol Genet Genomics 276:101–112

Cogan NOI, Drayton MC, Ponting RC, Vecchies AC, Bannan NR, Sawbridge TI, Smith KF, Spangenberg GC, Forster JW (2007) Validation of *in silico*-predicted genic SNPs in white clover (*Trifolium repens* L.), an outbreeding alloplyploid species. Mol Genet Genomics 277:413–425

Crane C, Wright E, Dixon RA, Wang Z-Y (2006) Transgenic *Medicago truncatula* plants obtained from *Agrobacterium tumefaciens*-transformed roots and *Agrobacterium rhizogenes*-transformed hairy roots. Planta 223:1344–1354

Curley J, Sim SC, Warnke S, Leong S, Barker R, Jung G (2005) QTL mapping of resistance to grey leaf spot in ryegrass. Theor Appl Genet 111:1107–1117

Dalton SJ, Bettany AJE, Timms E, Morris P (1999) Co-transformed, diploid *Lolium perenne* (Perennial ryegrass), *Lolium multiflorum* (Italian ryegrass) and *Lolium temulentum* (Darnel) plants produced by microprojectile bombardment. Plant Cell Rep 18:721–726

Denchev PD, Songstad DD, McDaniel JK, Conger BV (1997) Transgenic orchardgrass (*Dactylis glomerata*) plants by direct embryogenesis from microprojectile bombarded leaf cells. Plant Cell Rep 16:813–819

Diwan N, Bouton JH, Kochert G, Cregan PB (2000) Mapping of simple sequence repeat (SSR) DNA markers in diploid and tetraploid alfalfa. Theor Appl Genet 101:165–172

Dobrowolski MP, Forster JW (2007) Linkage disequilibrium-based association mapping in forage species. In: Oraguzie NC, Rikkerink E, Gardiner SE, De Silva NH (eds) Association mapping in plants. Springer, New York, pp 197–209

Dong S, Qu R (2005) High efficiency transformation of tall fescue with *Agrobacterium tumefaciens*. Plant Sci 168:1453–1458

Dumsday JL, Smith KF, Forster JW, Jones ES (2003) SSR-based genetic linkage analysis of resistance to crown rust (*Puccinia coronata* Corda f. sp. *lolii*) in perennial ryegrass (*Lolium perenne* L.). Plant Pathol 52:628–637

Ergon A, Fang C, Jorgensen O, Aamlid TS, Rognli OA (2006) Quantitative trait loci controlling vernalisation requirement, heading time and number of panicles in meadow fescue (*Festuca pratensis* Huds.). Theor Appl Genet 112:232–242

Gadegaard G, Didion T, Folling M, Storgaard M, Andersen CH, Nielsen KK (2008) Improved fructan accumulation in perennial ryegrass transformed with the onion fructosyltransferase genes 1-SST and 6G-FFT. J Plant Physiol 165:1214–1225

Gao C, Jiang L, Folling M, Han L, Nielsen KK (2006) Generation of large numbers of transgenic Kentucky bluegrass (*Poa pratensis* L.) plants following biolistic gene transfer. Plant Cell Rep 25:19–25

Ge Y, Norton T, Wang Z-Y (2006) Transgenic zoysiagrass (*Zoysia japonica*) plants obtainedby *Agrobacterium*-mediated transformation. Plant Cell Rep 25:792–798

Gill G P, Wilcox PL, Whittaker DJ, Winz RA, Bickerstaff P, Echt CE, Kent J, Humphreys MO, Elborough KM, Gardner RC (2006) A framework linkage map of perennial ryegrass based on SSR markers. Genome 49:354–364

Gondo T, Tsuruta S, Akashi R, Kawamura O, Hoffmann F (2005) Green herbicide-resistant plants by particle inflow gun-mediated gene transfer to diploid bahiagrass (*Paspalum notatum*). J Plant Physiol 162:1367–1375

Gondo T, Matsumoto J, Tsuruta S, Yoshida M, Kawakami A, Terami F, Ebina M, Yamada T, Akashi R (2009) Particle inflow gun-mediated transformation of multiple-shoot clumps in rhodes grass (*Chloris gayana*). J Plant Physiol 166:435–441

Guo Y-D, Hisano H, Shimamoto Y, Yamada T (2009) Transformation of androgenic-derived Festulolium plants (*Lolium perenne* L. x *Festuca pratensis* Huds.) by *Agrobacterium tumefaciens*. Plant Cell Tissue Organ Cult 96:219–227

Fang C, Aamlid TS, Jorgensen O, Rognli OA (2004) Phenotypic and genotypic variation in seed production traits within a full-sib family of meadow fescue. Plant Breed 123:241–246

Farrar K, Asp T, Lübberstedt T, Xu M, Thomas AM, Christiansen C, Humphreys MO, Donnison IS (2007) Construction of two *Lolium perenne* BAC libraries and identification of BACs containing candidate genes for disease resistance and forage quality. Mol Breed 19:15–23

Faville M, Vecchies AC, Schreiber M, Drayton MC, Hughes LJ, Jones ES, Guthridge KM, Smith KF, Sawbridge T, Spangenberg GC, Bryan GT, Forster JW (2004) Functionally-associated molecular genetic marker map construction in perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 110:12–32

Forster JW, Jones ES, Kölliker R, Drayton MC, Dumsday J, Dupal MP, Guthridge KM, Mahoney NL, van Zijll de Jong E, Smith KF (2001) Development and Implementation of Molecular Markers for Forage Crop Improvement. In: Spangenberg G (ed) Molecular breeding of forage crops. Kluwer Academic, Dordrecht, pp 101–133

Forster JW, Jones ES, Batley J, Smith KF (2004) Molecular marker-based genetic analysis of pasture and turf grasses. In: Hopkins A, Wang Z-Y, Sledge M, Barker RE (eds) Molecular breeding of Forage and Turf. Kluwer Academic, Dordrecht, pp 197–239

Francia E, Barabaschi D, Tondelli A, Laidò G, Rizza F, Stanca AM, Busconi M, Fogher C, Stockinger EJ, Pecchioni N (2007) Fine mapping of a *Hv CBF* gene cluster at the frost resistance locus *Fr-H2* in barley. Theor Appl Genet 115:1083–1091

Fredslund J, Madsen LH, Hougaard BK, Nielsen AM, Bertioli D, Sandal N, Stougaard J, Schauser L (2006) A general pipeline for the development of anchor markers for comparative genomics in plants. BMC Genomics 7:207

Ha CD, Lemaux PG, Cho MJ (2001) Stable transformation of a recalcitrant Kentucky bluegrass (*Poa pratensis* L.) cultivar using mature seed-derived highly regenerative tissues. In Vitro Cell Dev Biol - Plant 37:6–11

Hendry GAF (1993) Evolutionary origins and natural functions of fructans - a climatological, biogeographic and mechanistic appraisal. New Phytol 123:3–14

Herrmann D, Boller B, Studer B, Widmer F, Kölliker R (2006) QTL analysis of seed yield components in red clover (*Trifolium pratense* L.). Theor Appl Genet 112:536–545

Herrmann D, Boller B, Studer B, Widmer F, Kölliker R (2008) Improving persistence in red clover: insights from QTL analysis and comparative phenotypic evaluation. Crop Sci 48: 269–277

Hirata M, Cai H, Inoue M, Yuyama N, Miura Y, Komatsu T, Takamizo T, Fujimori M (2006) Development of simple sequence repeat (SSR) markers and construction of an SSR-based linkage map in Italian ryegrass (*Lolium multiflorum* Lam.). Theor Appl Genet 113:270–279

Hisano H, Kanazawa A, Kawakami A, Yoshida M, Shimamoto Y, Yamada T (2004) Transgenic perennial ryegrass plants expressing wheat fructosyltransferase genes accumulate increased amounts of fructan and acquire increased tolerance on a cellular level to freezing. Plant Sci 167:861–868

Hisano H, Kanazawa A, Yoshida M, Humphreys MO, Iizuka M, Kitamura K, Yamada T (2008) Coordinated expression of functionally diverse fructosyltransferase genes is associated with fructan accumulation in response to low temperature in perennial ryegrass. New Phytol 178:766–780

Ikeda S, Takahashi W, Oishi H (2004) Generation of expressed sequence tags from cDNA libraries of Italian ryegrass (*Lolium multiflorum* Lam.). Grassl Sci 49:593–598

Inoue M, Gao ZS, Hirata M, Fujimori M, Cai HW (2004a) Construction of a high-density map of Italian ryegrass (*Lolium multiflorum* Lam.) using restriction fragment length polymorphism, amplified fragment length polymorphism and telomeric repeat-associated sequence markers. Genome 47:57–65

Inoue M, Gao ZS, Cai HW (2004b) QTL analysis of lodging resistance and related traits in Italian ryegrass (*Lolium multiflorum* Lam.). Theor Appl Genet 109:1576–1585

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800

Ishikawa G, Yonemaru J, Saito M, Nakamura T (2007) PCR-based landmark unique gene (PLUG) markers effectively assign homoeologous wheat genes to A, B and D genomes. BMC Genomics 8:135

Isobe S, Klimenko I, Ivashuta S, Gau M, Kozlov NN (2003) First RFLP linkage map of red clover (*Trifolium pratense* L.) based on cDNA probes and its transferability to other red clover germplasm. Theor Appl Genet 108:105–112

Isobe S, Nakaya A, Tabata S (2007) Genotype matrix mapping: searching for quantitative trait loci interactions in genetic variation in complex traits. DNA Res 14:217–225

Jenkins CLD, Snow AJ, Simpson RJ, Higgins TJ, Jacques NA, Pritchard J, Gibson J, Larkin PJ (2002) Fructan formation in transgenic white clover expressing a fructosyltransferase from *Streptococcus salivarius*. Funct Plant Biol 29:1287–1298

Jensen LB, Andersen JR, Frei U, Xing Y, Taylor C, Holm PB, Lübberstedt T (2005) QTL mapping of vernalisation response in perennial ryegrass (*Lolium perenne* L.) reveals co-location with an orthologue of wheat *VRN1*. Theor Appl Genet 110:527–536

Jensen LB, Holm PB, Lübberstedt T (2007) Cross-species amplification of 105 *Lolium perenne* SSR loci in 23 species within the Poaceae. Mol Ecol Not 7:1155–1161

Jones ES, Dupal MP, Kölliker R, Drayton MC, Forster JW (2001) Development and characterisation of simple sequence repeat (SSR) markers for perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 102:405–415

Jones ES, Mahoney NL, Hayward MD, Armstead IP, Jones JG, Humphreys MO, King IP, Kishida T, Yamada T, Balfourier F, Charmet C, Forster JW (2002a) An enhanced molecular marker-based map of perennial ryegrass (*Lolium perenne* L.) reveals comparative relationships with other Poaceae species. Genome 45:282–295

Jones ES, Dupal MD, Dumsday JL, Hughes LJ, Forster JW (2002b) An SSR-based genetic linkage map for perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 105:577–584

Jones ES, Hughes LJ, Drayton MC, Abberton MT, Michaelson-Yeates TPT, Forster JW (2003) An SSR and AFLP molecular marker-based genetic map of white clover (*Trifolium repens* L.). Plant Sci 165:531–539

Kalla R, Chu P, Spangenberg G (2001) Moleular breeding of forage luguems for virus resistance. In: Spangenberg G (ed) Molecular breeding of forage and turf. Kluwer Academic, Dordrecht, pp 219–237

Kawakami A, Yoshida M (2002) Molecular characterization of sucrose:sucrose 1-fructosyltransferase and sucrose:fructan 6-fructosyltransferase associated with fructan accumulation in winter wheat during cold hardening. Biosci Biotechnol Biochem 66:2297–2305

Kölliker R, Jones ES, Drayton MC, Dupal MP, Forster JW (2001) Development and characterisation of simple sequence repeat (SSR) markers for white clover (*Trifolium repens* L.). Theor Appl Genet 102:416–424

Kubik C, Sawkins M, Meyer WA, Gaut BS (2001) Genetic diversity in seven perennial ryegrass (*Lolium perenne* L.) cultivars based on SSR markers. Crop Sci 41:1565–1572

Lasseur B, Lothier J, Djoumad A, De Coninck B, Smeekens S, Van Laere A, Morvan-Bertrand A, Van den Ende W, Prud'homme MP (2006) Molecular and functional characterization of a cDNA encoding fructan:fructan 6G-fructosyltransferase (6G-FFT)/ fructan:fructan 1-fructosyltransferase (1-FFT) from perennial ryegrass (*Lolium perenne* L.). J Exp Bot 57:2719–2734

Lee S-H, Lee D-G, Woo H-S, Lee K-W, Kim D-H, Kwak S-S, Kim J-S, Kim H, Ahsan N, Choi MS, Yang J-K, Lee B-H (2006) Production of transgenic orchardgrass via *Agrobacterium*-mediated transformation of seed-derived callus tissues. Plant Sci 171:408–414

Lem P, Lallemand J (2003) Grass consensus STS markers: an efficient approach for detecting polymorphism in *Lolium*. Theor Appl Genet 107:1113–1122

Li Q, Robson PRH, Bettany AJE, Donnison IS, Thomas H, Scott IM (2004) Modification of senescence in ryegrass transformed with IPT under the control of a monocot senescence-enhanced promoter. Plant Cell Rep 22:816–821

Mackie JM, Musial JM, Armour DJ, Phan HTT, Ellwood SE, Aitken KS, Irwin JAG (2007) Identification of QTL for reaction to three races of *Colletotrichum trifolii* and further analysis of inheritance of resistance in autotetraploid Lucerne. Theor Appl Genet 114:1417–1426

Miura Y, Ding CL, Ozaki R, Hirata M, Fujimori M, Takahashi W, Cai HW, Mizuno K (2005) Development of EST-derived CAPS and AFLP markers linked to a gene for resistance to ryegrass blast (*Pyricularia* sp.) in Italian ryegrass (*Lolium multiflorum* Lam.). Theor Appl Genet 111:811–818

Montague A, Ziauddin A, Lee R, Ainley WM, Strommer J (2007) High-efficiency phosphinothricin-based selection for alfalfa transformation. Plant Cell Tissue Organ Cult 91:29–36

Musial JM, Aitken KS, Mackie JM, Irwin JAG (2005) A genetic linkage map in autotetraploid lucerne adapted to northern Australia, and the use of the map to identify DNA markers linked to resistance to *Phytophthora medicaginis*. Aust J Agric Res 56:333–344

Musial JM, Aitken KS, Mackie JM, Irwin JAG (2006) DNA markers linked to yield, yield components, and morphological traits in autotetraploid lucerne (*Medicago sativa* L.). Aust J Agric Res 57:801–810

Musial JM, Mackie JM, Armour DJ, Phan HTT, Ellwood SE, Aitken KS, Irwin JAG (2007) Identification of QTL for resistance and susceptibility to *Stagonospora meliloti* in autotetraploid Lucerne. Theor Appl Genet 114:1427–1435

Muylle H, Baert J, Van Bockstaele E, Petijs J, Roldán-Ruiz I (2005a) Four QTLs determine crown rust (*Puccinia coronata* f.sp. *lolii*) resistance in a perennial ryegrass (*Lolium perenne*) population. Heredity 95:348–357

Muylle H, Baert J, Van Bockstaele E, Moerkerke B, Goetghebeur E, Roldán-Ruiz I (2005b) Identification of molecular markers linked with crown rust (*Puccinia coronata* f.sp. *lolii*) resistance in perennial ryegrass (*Lolium perenne*) using AFLP markers and a bulked segregant approach. Euphytica 143:135–144

Nakashima K, Yamaguchi-Shinozaki K (2006) Regulons inovolved in osmotic stress-responsive and cold stress-responsive gene expression in plants. Physiol Plant 126:62–71

Narasimhamoorthy B, Bouton JH, Olsen KM, Sledge MK (2007) Quantitative trait loci and candidate gene mapping of aluminum tolerance in diploid alfalfa. Theor Appl Genet 114:901–913

Pavis N, Boucaud J, Prud'homme MP (2001) Fructans and fructan-metabolizing enzymes in leaves of *Lolium perenne*. New Phytol 150:97–109

Petrovska N, Wu X, Donato R, Wang Z-Y, Ong E-K, Jones E, Forster J, Emmerling M, Sidoli A, O'Hehir R, Spangenberg G (2004) Transgenic ryegrasses (*Lolium* spp.) with down-regulation of main pollen allergens. Mol Breed 14:489–501

Ponting RC, Drayton MC, Cogan NOI, Dobrowolski MP, Spangenberg GC, Smith KF, Forster JW (2007) SNP discovery, validation, haplotype structure and linkage disequilibrium in full-length herbage nutritive quality genes of perennial ryegrass (*Lolium perenne* L.). Mol Genet Genomics 278:585–597

Richards HA, Rudas VA, Sun H, McDaniel JK, Tomaszewski Z, Conger BV (2001) Genetic transformation and hybridization: construction of a GFP-BAR plasmid and its use for switchgrass transformation. Plant Cell Rep 20:48–54

Robins JG, Luth D, Campbell TA, Bauchan GR, He C, Viands DR, Hansen JL, Brummer EC (2007a) Genetic mapping of biomass production in tetraploid alfalfa. Crop Sci 47:1–10

Robins JG, Bauchan GR, Brummer EC (2007b) Genetic mapping forage yield, plant height, and regrowth at multiple harvests in tetraploid alfalfa (*Medicago sativa* L.). Crop Sci 47:11–18

Rosellini R, Capomaccio S, Ferradini N, Sardaro MLS, Nicolia A, Veronesi F (2007) Non-antibiotic, efficient selection for alfalfa genetic engineering. Plant Cell Rep 26:1035–1044

Saha MC, Mian MAR, Eujayl I, Zwonitzer JC, Wang L, May GD (2004) Tall fescue EST-SSR markers with transferability across several grass species. Theor Appl Genet 109:783–791

Saha MC, Mian R, Zwonitzer JC, Chekhovskiy K, Hopkins AA (2005) An SSR- and AFLP-based genetic linkage map of tall fescue (Festuca *arundinacea* Schreb.). Theor Appl Genet 110:323–336

Saha MC, Cooper JD, Mian MAR, Chekhovskiy K, May GD (2006) Tall fescue genomic SSR markers: development and transferability across multiple grass species. Theor Appl Genet 113:1449–1458

Sanford JC (1988) The biolistic process. Trends Biotechnol 6:299–302
Sato H, Takamizo T (2006) *Agrobacterium tumefaciens*-mediated transformation of forage-type perennial ryegrass (*Lolium perenne* L.). Grassland Sci 52:95–98
Sato S, Isobe S, Asamizu E, Ohmido N, Kataoka R, Nakamura Y, Kaneko T, Sakurai N, Okumura K, Klimenko I, Sasamoto S, Wada T, Watanabe A, Kohara M, Fujishiro T, Tabata S (2005) Comprehensive structural analysis of the genome of red clover (*Trifolium pratense* L.). DNA Res 12:301–364
Sawbridge T, Ong E-K, Binnion C, Emmerling M, McInnes R, Meath K, Nguyen N, Nunan K, O'Neill M, O'Toole F, Rhodes C, Simmonds J, Tian P, Wearne K, Webster T, Winkworth A, Spangenberg G (2003a) Generation and analysis of expressed sequence tags in perennial ryegrass (*Lolium perenne* L.). Plant Sci 165:1089–1100
Sawbridge T, Ong E-K, Binnion C, Emmerling M, Meath K, Nunan K, O'Neill O, O'Toole F, Simmonds J, Wearne K, Winkworth A, Spangenberg G (2003b) Generation and analysis of expressed sequence tags in white clover (*Trifolium repens* L.). Plant Sci 165:1077–1089
Shinozuka H, Hisano H, Ponting RC, Jones ES, Cogan NOI, Forster JW, Yamada T (2005) Molecular cloning and genetic mapping of perennial ryegrass protein kinase CK2□-subunit genes. Theor Appl Genet 112:167–177
Shinozuka H, Hisano H, Yoneyama S, Shimamoto Y, Jones ES, Forster JW, Yamada T, Kanazawa A (2006) Cold-responsive gene expression, linkage mapping and protein localization in the nucleus of glycine-rich RNA-binding protein gene from perennial ryegrass suggest a role for cold adaptation. Mol Genet Genomics 275:399–408
Sim S, Chang T, Curley J, Warnke SE, Barker R, Jung G (2005) Chromosomal rearrangements differentiating the ryegrass genome from the Triticeae, oat and rice genomes using common heterologous RFLP probes. Theor Appl Genet 110:1011–1019
Sim S, Diesburg K, Casler M, Jung G (2007) Mapping and comparative analysis of QTL for crown rust resistance in an Italian x perennial ryegrass population. Phytopathology 97:767–776
Skinner JS, Zitzewitz von J, Szűcs P, Marquez-Cedillo L, Filichkin T, Amundsen K, Stockinger EJ, Thomashow MF, Chen THH, Hayes PM (2005) Structural, functional, and phylogenetic characterization of a large *CBF* gene family in barley. Plant Mol Biol 59:533–551
Skøt L, Humphreys J, Humphreys MO, Thorogood D, Gallagher J, Sanderson R, Armstead IP, Thomas ID (2007) Association of candidate genes with flowering time and water-soluble carbohydrate content in *Lolium perenne* L. Genetics 177:535–547
Sledge MK, Ray IM, Jiang G (2005) An expressed sequence tag SSR map of tetraploid alfalfa (*Medicago sativa* L.). Theor Appl Genet 111:980–992
Smith KF, Dobrowolski MP, Cogan NOI, Spangenberg GC, Forster JW (2008) Utilizing linkage disequilibrium and association mapping to implement candidate gene based markers in perennial ryegrass breeding. In: Yamada T, Spangenberg G (eds) Molecular breeding of forage and turf. Springer Science + Business Media, New York, pp 335–340
Somleva MN, Tomaszewski Z, Conger BV (2002) *Agrobacterium*-mediated genetic transformation of switchgrass. Crop Sci 42:2080–2087
Somleva MN, Snell KD, Beaulieu JJ, Peoples OP, Garrison BR, Patterson NA (2008) Production of polyhydroxybutyrate in switchgrass, a value-added co-product in an important lignocellulosic biomass crop. Plant Biotechnol J 6:663–678
Spangenberg G, Wang Z-Y, Wu XL, Nagel J, Iglesias VA, Potrykus I (1995a) Transgenic tall fescue (*Festuca arundinacea*) and red fescue (*F. rubra*) plants from microprojectile bombardment of embryogenic suspension cells. J Plant Physiol 145:693–701
Spangenberg G, Wang Z-Y, Wu XL, Nagel J, Potrykus I (1995b) Transgenic perennial ryegrass (*Lolium perenne*) plants from microprojectile bombardment of embryogenic suspension cells. Plant Sci 108:209–217
Spangenberg G, Kalla R, Lidgett A, Sawbridge T, Ong EK, John U (2001) Breeding forage plants in the genome era. In: Spangenberg G (ed) Molecular breeding of forage crops. Kluwer Academic, Dordrecht, pp 1–39
Spangenberg GS, Forster JW, Edwards D, John U, Mouradov A, Emmerling M, Batley J, Felitti S, Cogan NOI, Smith KF, Dobrowolski MP (2005) Future directions in the molecular breeding

of forage and turf. In: Humphreys MO (ed) Molecular breeding for the genetic improvement of forage crops and turf. Wageningen Academic, The Netherlands, pp 83–97

Studer B, Boller B, Herrmann D, Bauer E, Posselt UK, Widmer F, Kölliker R (2006) Genetic mapping reveals a single major QTL for bacterial wilt resistance in Italian ryegrass (*Lolium multiflorum* Lam.) Theor Appl Genet 113:66–671

Studer B, Boller B, Bauer E, Posselt UK, Widmer F, Kölliker R (2007) Consistent detection of QTLs for crown rust resistance in Italian ryegrass (*Lolium multiforum* Lam.) across environments and phenotyping methods. Theor Appl Genet 115:9–17

Studer B, Asp T, Frei U, Hentrup S, Meally H, Guillard A, Barth S, Muylle H, Rolda'n-Ruiz I, Barre P, Koning-Boucoiran C, Uenk-Stunnenberg G, Dolstra O, Skøt L, Skøt KP, Turner LB, Humphreys MO, Köllliker R, Roulund N, Nielsen KK, Lübberstedt T (2008) Expressed sequence tag-derived microsatellite markers of perennial ryegrass (*Lolium perenne* L.) Mol Breed 21:533–548

Takahashi W, Oishi H, Ebina M, Takamizo T, Komatsu T (2002) Production of transgenic Italian ryegrass (*Lolium multiflorum* Lam.) via microprojectile bomberdment of embryogenic calli. Plant Biotechnol 19:241–249

Takahashi W, Fujimori M, Miura Y, Komatsu T., Nishizawa Y, Hibi T, Takamizo T (2005) Increased resistance to crown rust disease in transgenic Italian ryegrass (*Lolium multiflorum* Lam.) expressing the rice chitinase gene. Plant Cell Rep 23:811–818

Tamura K, Yamada T (2007) A perennial ryegrass *CBF* gene cluster is located in a region predicted by conserved synteny between Poaceae species. Theor Appl Genet 114:273–283

Tamura T, Yonemaru J, Hisano H, Kanamori H, King J, King IP, Tase K, Sanada Y, Komatsu T, Yamada T (2009) Development of the intron-flanking EST markers for the *Lolium-Festuca* complex using rice genomic information. Theor Appl Genet 118:1549–1560

Thorogood D, Kaiser WJ, Jones JG, Armstead I (2002) Self-incompatibility in ryegrass 12: genotyping and mapping the S and Z loci of *Lolium perenne* L. Heredity 88:385–390

Tobias CM, Twigg P, Hayden DM, Vogel KP, Mitchell RM, Lazo GR, Chow EK, Sarath G (2005) Analysis of expressed sequence tags and the identification of associated short tandem repeats in switchgrass. Theor Appl Genet 111:956–964

Tobias CM, Hayden DM, Twigg P, Sarath G (2006) Genic microsatellite markers derived from EST sequencesof switchgrass (*Panicum virgatum* L.). Mol Ecol Not 6:185–187

Turner LB, Cairns AJ, Armstead IP, Ashton J, Skøt K, Whittaker D, Humphreys MO (2006) Dissecting the regulation of fructan metabolism in perennial ryegrass (*Lolium perenne*) with quantitative trait locus mapping. New Phytol 169:45–58

Turner LB, Cairns AJ, Armstead IP, Thomas H, Humphreys MW, Humphreys MO (2008) Does fructan have a functional role in physiological traits? Investigation by quantitative trait locus mapping. New Phytol 179:765–775

Wang Z-Y, Ge Y (2005a) *Agrobacterium*-mediated high efficiency transformation of tall fescue (*Festuca arundinacea* Schreb.). J Plant Physiol 162:103–113

Wang Z-Y, Ge Y (2005b) Rapid and efficient production of transgenic bermudagrass and creeping bentgrass bypassing the callus formation phase. Funct Plant Biol 32:769–776

Wang Z-Y, Yamada T (2008) Cool-season forage grasses. In: Kole C, Hall TC (eds) Compendium of transgenic crop plants: transgenic cereals and forage grasses. Wiley-Blackwell, Hoboken, NJ, pp 199–210

Wang Z-Y, Ye XD, Nagel J, Potrykus I, Spangenberg G (2001) Expression of a sulphur-rich sunflower albumin gene in transgenic tall fescue (*Festuca arundinacea* Schreb.) plants. Plant Cell Rep 20:213–219

Wang Z-Y, Bell J, Ge YX, Lehmann D (2003a) Inheritance of transgenes in transgenic tall fescue (*Festuca arundinacea* Schreb.). In Vitro Cell Dev Biol Plant 39:277–282

Wang Z-Y, Scott M, Bell J, Hopkins A, Lehmann D (2003b) Field performance of transgenic tall fescue (*Festuca arundinacea* Schreb.) plants and their progenies. Theor Appl Genet 107:406–412

Wang Z-Y, Lawrence R, Hopkins A, Bell J, Scott M (2004) Pollen-mediated transgene flow in the wind-pollinated grass species tall fescue (*Festuca arundinacea* Schreb.). Mol Breed 14:47–60

Wang Z-Y, Bell J, Cheng X, Ge Y, Ma X, Wright E, Xiao X, Zhang J, Bouton J (2008) Transgenesis in forage crops. In: Yamada T, Spangenberg G (eds) Molecular breeding of forage and turf. Springer Science + Business Media, New York, pp 335–340

Warnke SE, Barker RE, Jung G, Rouf Mian MA, Saha MC, Brilman LA, Dupal MD, Forster JW (2004) Genetic linkage mapping of an annual x perennial ryegrass population. Theor Appl Genet 109:294–304

Wright E, Dixon RA, Wang Z-Y (2006) *Medicago truncatula* transformation using cotyledon explants. In: Wang K (ed) *Agrobacterium* protocols, 2nd edn. Humana Press, Totowa, NJ, pp 129–135

Wu YY, Chen QJ, Chen M, Chen J, Wang XC (2005) Salt-tolerant transgenic perennial ryegrass (*Lolium perenne* L.) obtained by *Agrobacterium tumefaciens*-mediated transformation of the vacuolar Na+/H+ antiporter gene. Plant Sci 169:65–73

Wu YY, Chen QJ, Cui, XH, Chen H, Chen J, Wang XC (2007) Efficient regeneration and *Agrobacterium*-mediated stable transformation of perennial ryegrass. Russ J Plant Physiol 54:524–529

Xie D-Y, Sharma SB, Wright E, Wang Z-Y, Dixon RA (2006) Metabolic engineering of proanthocyanidins through co-expression of anthocyanidin reductase and the PAP1 MYB transcription factor. Plant J 45:895–907

Xiong Y, Fei S-Z, Brummer EC, Moore KJ, Barker RE, Jung G, Curley J, Warnke SE (2007a) QTL analyses of fiber components and crude protein in an annual x perennial ryegrass interspecific hybrid population. Mol Breed 18:327–340

Xiong Y, Fei S-Z, Arora R, Brummer EC, Barker RE, Jung G, Warnke SE (2007b) Identification of quantitative trait loci controlling winter hardiness in an annual x perennial ryegrass interspecific hybrid population. Mol Breed 19:125–136

Xu JP, Schubert J, Altpeter F (2001) Dissection of RNA-mediated ryegrass mosaic virus resistance in fertile transgenic perennial ryegrass (*Lolium perenne* L.). Plant J 26:265–274

Yamada T, Forster JW (2005) QTL analysis and trait dissection in ryegrasses (*Lolium* spp.). In: Humphreys MO (ed) Molecular breeding for the genetic improvement of forage crops and turf. Wageningen Academic, The Netherlands, pp 43–53

Yamada T, Jones ES, Cogan NOI, Vecchies AC, Nomura T, Hisano H, Shimamoto Y, Smith KF, Forster JW (2004) QTL analysis of morphological, developmental and winter hardiness-associated traits in perennial ryegrass (*Lolium perenne* L.). Crop Sci 44:925–935

Yamada T, Forster JW, Humphreys MW, Takamizo T (2005) Genetics and molecular breeding in *Lolium/Festuca* grass species complex. Grassl Sci 51:89–106

Yamaguchi-Shinozaki K, Shinozaki K (2006) Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stressed. Annu Rev Plant Biol 57:781–803

Yamaguchi-Shinozaki K, Shinozaki K (2008) DREB regulons in abiotic-stress-responsive gene expression in plants. In: Yamada T, Spangenberg G (eds) Molecular breeding of forage crops. Springer Science + Business Media, New York, pp 15–27

Ye X, Wang Z-Y, Wu X, Potrykus I, Spangenberg G (1997) Transgenic Italian ryegrass (*Lolium multiflorum*) plants from microprojectile bombardment of embryogenic suspension cells. Plant Cell Rep 16:379–384

Ye XD, Wu XL, Zhao H, Frehner M, Nosberger J, Potrykus I, Spangenberg G (2001) Altered fructan accumulation in transgenic *Lolium multiflorum* plants expressing a *Bacillus subtilis sacB* gene. Plant Cell Rep 20:205–212

Yoshida M, Abe J, Moriyama M, Kuwabara T (1998) Carbohydrate levels among wheat cultivars varying in freezing tolerance and snow mold resistance during autumn and winter. Physiol Plant 103:8–16

Zhang J-Y, Broeckling CD, Blancaflor EB, Sledge MK, Sumner LW, Wang Z-Y (2005) Overexpression of *WXPI*, a putative *Medicago truncatula* AP2 domain–containing transcription factor gene, increases cuticular wax accumulation and enhances drought tolerence in transgenic alfalfa (*Medicago sativa*). Plant J 42:689–707

Zhang H, Lomba P, Altpeter F (2007a) Improved turf quality of transgenic bahiagrass (*Paspalum notatum* Flugge) constitutively expressing the ATHB16 gene, a repressor of cell expansion. Mol Breed 20:415–423

Zhang L, Wu D, Zhang L, Yang C (2007b) *Agrobacterium*-mediated transformation of Japanese lawngrass (*Zoysia japonica* Steud.) containing a synthetic *cryIA(b)* gene from *Bacillus thuringiensis* Plant Breed 126:428–432

Zhang Y, Sledge MK, Bouton JH (2007c) Genome mapping of white clover (*Trifolium repens* L.) and comparative analysis within the Trifolieae using cross-species SSR markers. Theor Appl Genet 114:1367–1378

Author Index

B

D

H

K

M

N

O

Z